



“UN ANÁLISIS DE SENSIBILIDAD DE LA DESIGUALDAD DE INGRESOS EN EL PERÚ”

**Trabajo de Investigación presentado
para optar al Grado Académico de
Magíster en Economía**

**Presentado por
Bruno Mauricio Escobar Izquierdo**

Asesor: Diego Martín Winkelried Quezada

[0000-0002-9388-2617](tel:0000-0002-9388-2617)

Lima, setiembre 2020

Resumen Ejecutivo

Es un hecho estilizado conocido que el Perú, junto a otros países de América Latina, ha registrado una fuerte caída en la desigualdad de ingresos en las primeras décadas del siglo XXI. La llamada literatura de *Top Incomes* argumentaría que estas estadísticas oficiales estarían subestimadas, pues emplean únicamente datos de encuestas de hogares y estas típicamente no ofrecen una buena representación de los hogares con mayores ingresos. En este trabajo, buscamos atender algunas de estas críticas al uso de encuestas de hogares para cuantificar la desigualdad de ingresos en el Perú. En primer lugar, aplicamos una corrección semiparamétrica a los datos de ingresos totales de la Encuesta Nacional de Hogares (Enaho) para lidiar específicamente con el problema de cobertura dispersa de los hogares con mayores ingresos de la población. En segundo lugar, para abordar el problema más severo de cobertura truncada, realizamos diversas simulaciones de Montecarlo del coeficiente de Gini para diversos supuestos de la cola superior de la distribución de ingresos a lo largo del tiempo. Nuestras correcciones no producen cambios significativos del coeficiente de Gini, aunque deben interpretarse como cotas inferiores del verdadero coeficiente. Asimismo, nuestras simulaciones muestran que los supuestos necesarios para corregir al alza el coeficiente de Gini están dentro de los niveles observados en otros países, aunque no lo corrigen en la magnitud encontrada en trabajos previos. Más aun, los supuestos necesarios para rechazar la caída del Gini entre el 2004 y el 2018 serían improbables, dadas las estadísticas corregidas en otros países, con lo cual este hecho estilizado sería robusto a diversas críticas en las encuestas de hogares.

Palabras Clave: *Top Incomes*, encuestas de hogares, subcobertura, desigualdad de ingresos, Perú, coeficiente de Gini, simulación de Montecarlo.

Código JEL: D63, O54, F63, N36

Índice de Contenidos

Índice de Figuras.....	v
Índice de Apéndices	vi
 Capítulo I. Introducción.....	 1
 Capítulo II. Problemas de Muestreo en las Encuestas de Hogares	 8
1. Problemas de Cobertura Dispersa.....	8
2. Problemas de Cobertura Truncada.....	10
3. WID.World: Evidencia Internacional de la Desigualdad Corregida	13
 Capítulo III. Discusión Metodológica	 18
1. Modelo de la Distribución de Ingresos	18
2. Corrección Semiparamétrica del Coeficiente de Gini	23
3. Algoritmo de Simulación de Montecarlo del Coeficiente de Gini	25
4. Limitaciones.....	27
 Capítulo IV. Datos y Estadísticos	 29
1. Descripción de las Bases de Datos.....	29
2. Estadísticos Descriptivos de la Desigualdad en el Perú.....	30
3. Evidencia de Correcciones al Coeficiente de Gini en el Perú	34
 Capítulo V. Análisis de las Correcciones	 36

1. Estimaciones del Índice de la Cola	36
2. Estimaciones de las Correcciones por Cobertura Dispersa.....	40
 Capítulo VI. Simulaciones del Coeficiente de Gini	43
1. Simulaciones de las Correcciones Anuales del Coeficiente de Gini	43
2. Simulaciones del Cambio del Coeficiente de Gini en el Tiempo	47
3. ¿Qué nos Dicen las Simulaciones sobre el Gini del Perú?	53
 Conclusiones y Recomendaciones.....	55
 Referencias.....	57
 Anexos	64

Índice de Figuras

Figura 1.	Participación del 1 % más rico en diversos países del mundo	15
Figura 2.	Evolución de estadísticos de desigualdad de ingresos en el Perú: 1997 a 2018	32
Figura 3.	Serie de tiempo de las estimaciones del índice de la cola superior del ingreso para diferentes clasificaciones de hogares con mayores ingresos.....	38
Figura 4.	Estimaciones del índice de la cola superior de ingresos para diferentes años de la Enaho y diferentes proporciones de hogares no observados.....	39
Figura 5.	Serie de tiempo de estadísticas de la desigualdad de ingresos, con y sin correcciones	40
Figura 6.	Corte transversal del coeficiente de Gini corregido	41
Figura 7.	Probabilidades simuladas de observar una corrección positiva en el coeficiente de Gini.....	44
Figura 8.	Probabilidades simuladas de observar correcciones de diversa magnitud en el coeficiente de Gini del año 2012.....	47
Figura 9.	Probabilidades simuladas de observar una reducción del coeficiente de Gini entre 2004 y 2018.....	49
Figura 10.	Probabilidades simuladas de observar una reducción del coeficiente de Gini por subperíodos	50
Figura 11.	Trayectorias simuladas del coeficiente de Gini bajo supuestos comunes.....	52

Índice de Anexos

Tabla A1. Estimaciones del índice de la cola superior de ingresos en Perú: 2004-2018	64
Figura A1. Participación del 10 % más rico en diversos países del mundo	66
Figura A2. Estadísticas corregidas de la desigualdad del ingreso neto anual.....	66
Figura A3. Estadísticas corregidas de la desigualdad del ingreso neto equivalente de los hogares	67
Figura A4. Probabilidades simuladas de observar correcciones de diversa magnitud en el coeficiente de Gini del año 2004.....	68
Figura A5. Probabilidades simuladas de observar correcciones de diversa magnitud en el coeficiente de Gini del año 2018.....	69
Figura A6. Probabilidades simuladas de observar correcciones de diversa magnitud en el coeficiente de Gini del año 1997.....	70
Figura A7. Probabilidades simuladas de observar cambios del Gini del ingreso bruto anual per cápita para varios años	71

Capítulo I. Introducción

Es un hecho estilizado bien conocido que la primera década del siglo XXI marcó un punto de inflexión en la tendencia de la desigualdad de ingresos en América Latina (Gasparini, Cruces y Tornarolli, 2011; Lustig, Lopez-Calva y Ortiz-Juarez, 2013). De tener un crecimiento de la desigualdad en la década de 1990, numerosos estudios y estadísticas oficiales reportan que la mayoría de países en la región reportaron una significativa caída del coeficiente de Gini de los ingresos entre los primeros años de los 2000 y los años posteriores a 2010. Esta evolución en la desigualdad es particularmente importante, porque vino de la mano de un periodo de numerosos desarrollos positivos en la región (Lustig, Lopez-Calva y Ortiz-Juarez, 2013; Gasparini, Cruces y Tornarolli, 2011; Amarante, 2016; Szekely y Mendoza, 2017; Cepal, 2019; Ferreira *et al.*, 2013). Entre estos países, el Perú es comparativamente uno de los casos de mayor éxito, pues reportó grandes reducciones de pobreza y alta movilidad social acompañando al ciclo económico (Ferreira *et al.*, 2013; Genoni y Salazar, 2015; Winkelried y Torres, 2019).

Frente a estos resultados notables, existe una amplia literatura que criticaría las estadísticas sobre la desigualdad, debido a que emplean exclusivamente datos de las encuestas de hogares (Atkinson, Piketty y Saez, 2011; Burkhauser, Herauld, Wilkins y Jenkins, 2016; Jenkins, 2017; Lakner y Milanovic, 2016; Vermuelen, 2018; Piketty y Saez, 2003; Bach, Corneo y Steiner, 2009; Anand y Segal, 2017). Estos estudios acerca de la literatura de la desigualdad de ingresos argumentan que las encuestas de hogares sufren de múltiples problemas de subcobertura de los hogares con los ingresos más altos de la distribución y, por ello, los estadísticos oficiales de desigualdad estarían subestimados. Estas críticas son válidas, y nos llevan a preguntar si se mantendría el tamaño y la tendencia histórica de las cifras oficiales en caso de que tuviésemos

acceso a datos representativos de los hogares más ricos en el Perú para complementar la Encuesta Nacional de Hogares (Enaho).

Buscamos atender directamente dos modalidades del problema de representatividad de las encuestas de hogares bajo un marco conceptual unificado. Una modalidad específica que se puede corregir utilizando únicamente datos de encuestas de hogares es el problema de cobertura dispersa, según el cual la encuesta arroja una muestra de hogares ricos con altos niveles de dispersión entre ellos, y esto resulta en estadísticos con baja precisión y potencialmente sesgados (Jenkins, 2017; Lustig, 2019). En estos casos es posible corregir el sesgo inducido en los índices de desigualdad aplicando el método semiparamétrico propuesto en Cowell y Flachaire (2007) y Davidson y Flachaire (2007). El método requiere estimar una distribución paramétrica conocida a los ingresos de una proporción fija de los hogares más ricos de la población. La distribución elegida en la práctica ha sido predominantemente la distribución de Pareto (Jenkins, 2017; Charpentier y Flachaire, 2019), mientras que el umbral que separa a los hogares ricos y el resto de la población es elegido por el investigador. Luego, se busca reemplazar el ingreso de la muestra original de hogares ricos con realizaciones aleatorias de la distribución estimada y recalcular el coeficiente de Gini. El Gini calculado con la nueva muestra ya no exhibirá los sesgos propios de una muestra con ingresos altos esparcidos ni *outliers*, pues por construcción habrá densidad de masa positiva en cualquier segmento del soporte de la distribución.

En una modalidad más severa del problema, una fracción de los hogares más ricos en la población jamás es muestreada y, por tanto, se observa una muestra que sufre de cobertura truncada (por la derecha). Las correcciones basadas únicamente en datos de encuestas de hogares tienen una capacidad limitada para subsanar problema, así que típicamente las correcciones complementan los datos de encuestas de hogares con datos de fuentes externas que capturen

mejor los ingresos de los sectores más ricos de la población para obtener. Los datos de impuestos nacionales han sido la fuente externa que mayor atención ha recibido para ello (Atkinson, 2007).

Una primera dificultad que enfrentan el Perú y la mayoría de países de América Latina es que no hay acceso público a los registros de los impuestos de la población, inclusive tabulaciones, y solo se ha podido aplicar correcciones basadas en imputación por cuentas nacionales, obteniendo correcciones de diversos tamaños pero, en general, acordando que el Gini de la Enaho estaría subestimado (Yamada, Castro y Bacigalupo, 2012; Cruz-Saco, Seminario y Campos, 2018; Yamada y Castro, 2006; Yamada, Castro y Oviedo, 2016; Mendoza, Leyva y Flor, 2011). Pero incluso si se contase con la mayor calidad de datos disponibles de diversas fuentes, queda una incertidumbre de especificación fundamental: ¿Cómo se combinan estos datos para hallar la distribución verdadera? Más importante aun, resulta que las diversas decisiones que el investigador debe hacer para obtener una corrección pueden tener efectos importantes sobre la corrección final obtenida (Lustig, 2019).

Abordamos este problema realizando numerosas simulaciones de Montecarlo del coeficiente de Gini en las cuales —de forma similar a las correcciones— reemplazamos los datos de ingresos de hogares ricos con pseudodatos de una distribución sobre la cual controlamos y alteramos sus parámetros. Buscamos identificar numéricamente los supuestos paramétricos que uno debería creer para obtener una corrección positiva del coeficiente de Gini (medido como una probabilidad menor a 0.9 de que el Gini simulado en un año sea mayor al de su muestra original), y rechazar la caída de la desigualdad en el Perú entre el 2004 y el 2018 (medido como una probabilidad menor a 0.9 de que el Gini simulado en 2004 sea mayor al de 2018). Específicamente, los supuestos que variamos son la proporción de la distribución no observada, y un único parámetro que gobierna esta distribución no observada. Luego, comparamos estos

parámetros con sus análogos empíricos en otros países donde sí se han empleado correcciones completas para establecer si son supuestos creíbles para asumir.

Usamos la Enaho del Perú entre 1997 y 2018. Las estadísticas descriptivas rápidas verifican que el coeficiente de Gini cayó significativamente en este período, posiblemente debido a una caída en la proporción del ingreso de los percentiles más altos de la población. Sin embargo, este desarrollo del coeficiente sería, en todo caso, representativo de la evolución de la desigualdad de los ingresos laborales (Székely y Hilgert, 2007; Alarco, Castillo y Leiva, 2019), lo cual motiva la aplicación de nuestros dos métodos. Por un lado, nuestra serie corregida del coeficiente de Gini no muestra ningún cambio significativo ni en precisión, sugiriendo que el coeficiente oficial es una cota inferior apropiada del verdadero coeficiente. Esto se mantiene para cualquier clasificación de hogares ricos que empleamos, y para las definiciones alternativas de la variable de análisis. Por otro lado, las simulaciones muestran que los supuestos necesarios para incrementar de forma significativa el coeficiente de Gini están dentro de los niveles observados en otros países, aunque no corrigen el coeficiente en la magnitud hallada en otros trabajos. Además, los supuestos necesarios para rechazar la caída del Gini entre el 2004 y el 2018 son más extremos e improbables dadas las estadísticas corregidas en otros países, en tanto sugieren que el Perú tendría niveles de concentración de ingreso al menos tan altos como los países más desiguales del mundo.

El trabajo ofrece cuatro contribuciones importantes. En primer lugar, nuestras correcciones aportan a la literatura empírica de la desigualdad de ingresos con información valiosa sobre el caso peruano. Este caso siempre ha sido un reto para analizar con los métodos clásicos que involucran complementar los datos de encuestas de hogares, aun con tabulaciones de impuestos, pues no hay acceso público a los datos de impuestos por la autoridad local. En

particular, las correcciones están más alineadas a aquellos trabajos que buscan corregir el coeficiente de Gini solo con encuestas de hogares. Estos incluyen a Alfons, Templ y Filzmoser (2013) y a Ruiz y Woloszko (2016); y particularmente a Cowell y Flachaire (2007) y a Davidson y Flachaire (2007), pues proveemos las primeras aplicaciones empíricas en América Latina de sus métodos. Los métodos del trabajo recaen enteramente dentro de la metodología de reemplazo de datos catalogada por Lustig (2019).

En segundo lugar, este es uno de los pocos estudios que emplean simulaciones de Montecarlo para estudiar la desigualdad como respuesta a la falta de datos representativos de los hogares ricos (Higgins, Lustig y Vigorito, 2018; Alarco, Castillo y Leiva, 2019). Hasta donde sabemos, este sería el primer trabajo que realiza simulaciones específicamente para cuantificar el cambio del coeficiente de Gini de un país a lo largo el tiempo, y mapear los parámetros (supuestos) de la simulación hacia probabilidades de observar cambios negativos (o positivos) en el tiempo.

En tercer lugar, a través de nuestras simulaciones también nos conectamos con el grueso de la literatura de *Top Incomes* que emplea fuentes adicionales a las encuestas de hogares. Las simulaciones exponen el rol crítico que juegan los supuestos que impone el investigador sobre la distribución de ingresos en todas las correcciones, y sugieren, tal y como intuye Lustig (2019), que toda corrección debería presentarse en rangos (en vez de puntos) y con una discusión completa de los supuestos subyacentes y sus implicancias.

La cuarta contribución es el mensaje de *policy-making*. Nuestras correcciones y simulaciones indican que, bajo una diversidad de escenarios sensatos, la caída en la desigualdad de ingresos desde el año 2004 se mantiene como un hecho estilizado en el cual los hacedores de política peruanos pueden confiar. En vista de estos resultados para el caso peruano, debería

resultar atractivo emplear nuestros métodos en otros países de la región, más aun dada la coyuntura actual. Tan solo el año pasado (2019), en Chile, país que ya es miembro de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), hubo una ola masiva de protestas violentas basadas en una sensación de descontento con respecto a la desigualdad percibida en la población¹. La preocupación por la desigualdad en otros países no se aleja mucho, es particularmente alta en Brasil (Morgan, 2018), y ya antes del 2000 la región registraba algunos de los mayores índices de desigualdad en el mundo (Lustig, Lopez-Calva y Ortiz-Juarez, 2013; Genoni y Salazar, 2015). Nuestros ejercicios informan al hacedor de política sobre la sensibilidad de la desigualdad de ingresos a los supuestos sobre los datos no observados en la encuesta de hogares local.

La distribución del trabajo es la siguiente. El capítulo II resume los principales problemas de muestreo asociados con las encuestas de hogares, las opciones para corregirlos, sus aplicaciones en otros países e, importantemente, sus limitaciones y nuestras opiniones sobre ellos. El Capítulo III ofrece una discusión metodológica donde se formaliza el modelo de ingresos subyacente, el método de corrección del coeficiente de Gini y el algoritmo para las simulaciones. También se discuten las limitaciones asociadas a nuestros métodos. En el Capítulo IV se describen los datos que se usarán, se presentan algunos estadísticos descriptivos de la desigualdad de ingresos en el Perú y se evalúa la evidencia de las correcciones realizadas hasta el momento para este país. En el Capítulo V se presentan los resultados de las correcciones del coeficiente de Gini. En el Capítulo VI se analizan los resultados de las simulaciones y se contrasta con los estadísticos disponibles en la base de WID.World. Finalmente, presentamos las

¹ Un artículo de The Economist (2019) resume los puntos clave de este reciente episodio. La preocupación es validada por estudios académicos también: El reciente trabajo de Flores *et al.* (2019) también estima altos niveles de concentración de ingresos en los percentiles más altos de la población.

conclusiones y proponemos recomendaciones de política y oportunidades futuras de investigación.

Capítulo II. Problemas de Muestreo en las Encuestas de Hogares

En la literatura de *Top Incomes* se han identificado y estudiado múltiples fuentes de subcobertura de hogares ricos que pueden aquejar a las encuestas de hogares. Siguiendo de cerca la exposición de Jenkins (2017) y Lustig (2019), clasificamos la mayoría de estos problemas en dos grupos: problemas que resultan en cobertura dispersa y aquellos que resultan en cobertura truncada². Utilizamos esta clasificación porque cada categoría sugiere una forma de corrección diferente de la otra. La diferencia no radica fundamentalmente en el método, sino en el tipo de datos que se requieren para subsanar las deficiencias de las encuestas de hogares. En esta sección, explicamos cada grupo, los métodos más populares de corrección propuestos con sus limitaciones, y las aplicaciones empíricas en otros países que resultan más relevantes para nuestro estudio.

1. Problemas de Cobertura Dispersa

Bajo el problema de cobertura dispersa, la encuesta arroja una muestra de hogares ricos con altos niveles de dispersión entre ellos. En este marco, los datos observados de los hogares más ricos sí provienen de una distribución común, pero el muestreo de estos fue pobre, en el sentido de que no fue completamente aleatorio, y la muestra exhibe tramos en la cola superior donde la densidad es nula. Esto puede suceder por varios motivos. Si la probabilidad de muestrear hogares ultrarricos es positiva pero casi nula, la frecuencia de *outliers* es alta. Asimismo, si los hogares ricos están subrepresentados desde el diseño de la muestra, o si tienden a subreportar sus ingresos, los datos pueden exhibir estos huecos artificiales.

² Existe una variante adicional que corresponde a *problemas de censura* (por la derecha) en la codificación de ingresos (Burkhauser, Feng, Jenkins y Larrimore, 2012). Dado que la censura no es un problema documentado en las encuestas peruanas, no profundizamos en más detalle sobre esta rama extensa de la literatura.

En todos estos casos prevalece la idea de que la encuesta de hogares sí ofrece observaciones a lo largo de todo el soporte de la distribución y, por tanto, las encuestas serían autosuficientes para corregir las deficiencias de los estimadores de desigualdad. Bajo esta premisa se desprenden los criterios principales para corregir la dispersión de datos y la presencia de *outliers*, propuestos por Davidson y Flachaire (2007) y Cowell y Flachaire (2007).(; Estos trabajos asumen que la cola superior de ingresos sigue una distribución conocida y, más crucialmente, que los parámetros de esta distribución se pueden estimar consistentemente solo con datos de encuestas de hogares. Luego, es válido recalcular la desigualdad reemplazando las observaciones de ingreso de los hogares ricos con pseudodatos, tomados aleatoriamente de la distribución estimada.

Este método resulta particularmente útil en el contexto peruano, en tanto tiene la menor exigencia de datos adicionales de todos los métodos, y porque es muy fácil de aplicar, así que lo implementamos. No obstante, debemos remarcar limitaciones importantes en el método. Ruiz y Woloszko (2016) presentan una estimación de la distribución de la cola superior de ingresos que, ellos aseguran, corrige el subreporte de ingresos de los ricos tan solo usando datos de encuestas de hogares. Somos escépticos de esta afirmación, pues otros trabajos evidencian que es inevitable subestimar los parámetros de la verdadera cola superior de ingresos con extrapolaciones basadas exclusivamente en datos de encuestas de hogares (Burkhauser, Feng, Jenkins y Larrimore, 2012; Atkinson, Piketty y Saez, 2011; Higgins, Lustig y Vigorito, 2018). Además, si bien se ha documentado que la dispersión de datos disminuye la precisión de los estimados no paramétricos de la desigualdad (Cowell y Flachaire, 2007), el signo del sesgo en los estimados atribuido a este problema no es conocido inmediatamente. Si la dispersión de datos se debe a una subrepresentación o subreporte del ingreso de hogares ricos en la muestra, podría

haberse subestimado la desigualdad (Jenkins, 2017), pero si la dispersión es resultado de un error de muestreo en una muestra perfectamente diseñada, no debería haber sesgo (Lustig, 2019), e incluso existe la posibilidad de que los *outliers* conlleven a sobreestimar la desigualdad (Hubert, Dierckx y Vanpaemel, 2013).

2. Problemas de Cobertura Truncada

Bajo el problema de cobertura truncada, el submuestreo de hogares ricos es de tal magnitud que una fracción de los hogares más ricos en la población jamás es muestreada en la encuesta de hogares, y la muestra observada en la práctica estaría truncada (por la derecha). Las circunstancias que pueden llevar a esta situación se pueden pensar como casos más severos de los factores que conducen a cobertura dispersa. Siguiendo la nomenclatura de Lustig (2019), en estos casos definitivamente existe subreporte y la subrepresentación de hogares de la verdadera cola superior de ingresos y, además, vienen usualmente acompañados de *item non-response* (cuando los individuos se niegan a responder ítems de la encuesta que competen a su ingreso) y *unit non-response* (cuando un individuo u hogar entero no es encuestado). Entonces, el problema es la ausencia de cobertura: existe una proporción de los hogares más ricos en la población cuya probabilidad de ser encuestada es nula.

La recomendación metodológica para lidiar con este problema, en principio, es bastante directa: Debe usarse una fuente de datos externa que capture mejor a los hogares con mayores ingresos. Es así como trabajos tempranos reconsideran por completo la distribución de ingresos y reestiman todos los cálculos de desigualdad usando exclusivamente datos de los registros históricos de impuestos. Esto se hizo primero en Francia, los Estados Unidos y el Reino Unido, hallando efectivamente incrementos considerables en la desigualdad (Piketty y Saez, 2003; Atkinson, 2005; Piketty, 2003).

Si bien estos trabajos han sido muy influyentes, posteriormente la literatura encontró que, debido a los problemas de representatividad de los datos de impuestos en la cola inferior de la distribución, el uso más apropiado de la fuente externa sería describir únicamente el ingreso de los hogares más ricos de la población (Lustig, 2019). Por su parte, las encuestas de hogares describen correctamente el ingreso del resto de la población. Luego, y si los grupos no se traslapan, diversos estadísticos poblacionales, como el coeficiente de Gini, en muchos casos son fórmulas conocidas de estadísticos de ambos grupos (Atkinson, 2007; Alvaredo, 2011). Este enfoque encontró muchas aplicaciones en países desarrollados con microdatos de impuestos disponibles. Los casos más estudiados son, nuevamente, el Reino Unido (Burkhauser, Herault, Wilkins y Jenkins, 2016; Jenkins, 2017; Atkinson y Piketty, 2010) y los Estados Unidos (Atkinson, Piketty y Saez, 2011; Alvaredo, 2011), donde resaltan que la desigualdad ha sido subestimada y la suave tendencia creciente de los estadísticos oficiales en estos países es, en realidad, mucho más pronunciada.

Lamentablemente, en pocos países se cuenta con acceso público a microdatos de las declaraciones de impuestos. Ante ello, existen varias alternativas para complementar los datos de encuestas de hogares. La propuesta alternativa que tiene mayor grado de aplicabilidad³ involucra el uso de la información macroeconómica de las cuentas nacionales. En su forma más simple, propone imputar el diferencial entre el ingreso per cápita de las cuentas nacionales y la media de ingresos en encuesta de hogar a la participación en el ingreso de los percentiles más altos de la distribución de ingresos, que no se observan en la encuesta de hogares. El método les ha

³ Otra solución innovadora consiste en recurrir a las listas de Forbes de los individuos más ricos del mundo por país (Forbes' World Billionaires) para imputar el extremo derecho de la cola de la distribución de la riqueza. Vermuelen (2018) aplica esta idea a los Estados Unidos, el Reino Unido y nueve países de la Eurozona con gran éxito, pues replican los resultados cualitativos de los trabajos basados en datos de impuestos. Su aplicabilidad se reduce notablemente para los países en desarrollo, y no ofrece datos de la distribución de ingresos.

permitido a Lakner y Milanovic (2016) y a Anand y Segal (2017) revisar los coeficientes de Gini en países de todo el mundo. El problema es que el método asume que el total de la diferencia entre el ingreso de cuentas nacionales y el ingreso de encuestas de hogares se pueda atribuir únicamente al ingreso de los hogares no observados. Sobre este diferencial tenemos poco entendimiento. Podría haber un problema de sumar “manzanas + plátanos” en la comparabilidad de dos fuentes de datos tan dispares (Jenkins, 2017). Además, como las cuentas nacionales están sujetas a errores de medición y correcciones continuas, en esta aplicación, resultarían en cambios importantes y artificiales en el Gini estimado (Deaton, 2005; Bourguignon, 2015). En todo caso, esta familia de correcciones tendería a sobreestimar el Gini verdadero.

En una aplicación más sofisticada del método de cuentas nacionales, popularizada en el reconocido trabajo de Piketty, Saez y Zucman (2018), se puede combinar los datos macroeconómicos, datos de encuestas de hogares y datos de registros de impuestos —que pueden ser solo tabulaciones de impuestos en vez de microdatos, y que muchos más países sí están dispuestos a publicar— para obtener una estimación completa de la distribución de ingresos del 100 % de la población que calce con el desarrollo macroeconómico histórico del país.

En el caso del Perú, no pretendemos aplicar metodologías de impuestos, ni otra variante del método de cuentas nacionales. Esto no se debe solamente al hecho de que aún no se cuenta con tabulaciones públicas de los registros de impuestos para implementar las correcciones, sino también a un problema de fondo. Aunque el método DINA sea el más completo disponible, esta sofisticación conlleva el costo de una larga serie de supuestos y operaciones necesarios para empalmar los datos de tres fuentes de datos tan diferentes. Al respecto, Lustig (2019) documenta que el dramático incremento de la desigualdad reportado en el influyente trabajo de Piketty, Saez

y Zucman (2018) ha sido cuestionado duramente por otros investigadores que emplean la misma metodología DINA en los Estados Unidos, pero que alteran la manipulación de los datos de impuestos y los supuestos de imputación y obtienen cifras mucho más moderadas de la desigualdad.

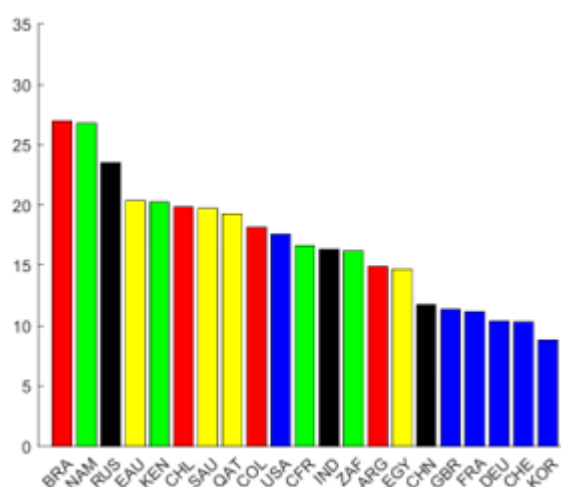
Uno de los objetivos de nuestra propuesta para analizar este problema, para el caso peruano, es justamente analizar la dependencia de las estimaciones de desigualdad a los supuestos que el investigador aplica en sus correcciones. Planteamos usar simulaciones de Montecarlo que nos permitan mapear dos supuestos simples sobre el tramo no observado (qué proporción no se observa y qué parámetro gobierna la distribución de esta parte no observada) hacia diversos estadísticos de interés, como la probabilidad de observar una caída del Gini entre dos momentos en el tiempo. En la literatura, los ejercicios de simulación han recibido poca atención, con la mención notable de Higgins, Lustig y Vigorito (2018) que usan simulaciones para confirmar que una submuestra de hogares uruguayos subreportan sus ingresos en comparación a su declaración de impuestos, y Alarco, Castillo y Leiva (2019) que implementan simulaciones a las correcciones basadas en el ajuste con las cuentas nacionales del Perú. Al margen de estos trabajos, no se ha empleado antes simulaciones para evaluar directamente la probabilidad de cambio del Gini de un país en el tiempo, bajo un modelo con cola de Pareto, y cuyo énfasis sea en el mapeo de supuestos paramétricos hacia probabilidades de eventos.

3. WID.World: Evidencia Internacional de la Desigualdad Corregida

Con todo, las correcciones que se han realizado a lo largo del mundo nos serán muy útiles. En años recientes, los métodos basados en impuestos han sido extensamente aplicados en todo el mundo debido a la iniciativa del World Inequality Database (WID) World Database, que busca recalcular la desigualdad en prácticamente todas las regiones del mundo. Esto incluye al

Medio Oriente (Alvaredo, Assouad y Piketty, 2019), África (Chancel, Cogneau, Gethin y Myczkowski, 2019), toda Europa (Blanchet, Chancel y Gethin, 2019) y varios países de América Latina que ya disponen de tabulaciones agregadas de datos de impuestos como Argentina (Alvaredo, 2011), Colombia (Alvaredo y Londoño-Vélez, 2013; Díaz-Bazán, 2015), Chile (Flores, Sanhueza, Atria, y Mayer, 2019) y Brasil (Morgan, 2018; Assouad, Chancel y Morgan, 2018). Otros países incluyen China (Piketty, Yang y Zucman, 2019), India (Chancel y Piketty, 2019), Corea del Sur (Kim, 2018), y Rusia (Novokmet, Piketty y Zucman, 2018). En la Figura 1 presentamos la participación en el ingreso de diversos países del mundo corregida principalmente por diversos métodos de impuestos que están recopiladas y disponibles para todo el público en la base de datos del WID.World. Estas cifras servirán de *benchmark* para poner en perspectiva el caso del Perú cuando interpretemos las simulaciones, pues nos permitirán establecer qué niveles de participación de ingresos serían sensatos en relación con las cifras de otros países.

(a) Promedio de la participación en el ingreso nacional del 1 % más rico, 2000 - 2003



(b) Promedio de la participación en el ingreso nacional del 1 % más rico, 2015 - 2018

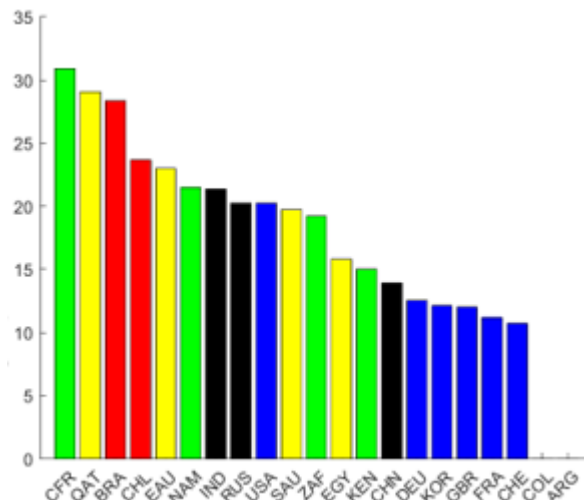


Figura 1. Participación del 1 % más rico en diversos países del mundo. Adaptado de WID.world.

Notas: Elegimos algunos países de la base de WID.World; sus nombres son abreviados a su código internacional de 3 letras. Las barras de países desarrollados (miembros de la OECD) están rellenas en azul, y los demás países según región: América Latina (rojo), Asia (negro), Medio Oriente (amarillo) y África (verde). Todas las estadísticas han sido corregidas según alguna variante de la metodología DINA (con diferentes niveles de sofisticación, dependiendo del acceso a datos en ese país); los estudios de donde se WID.World recopilan estos datos están citados en la discusión de la metodología DINA de la sección II. En general, la definición de ingreso corresponde al ingreso nacional repartido a nivel de individuos, en vez de hogares, y antes de impuestos.

Podemos observar que la participación del 1 % con mayores ingresos en América Latina no pasaba del 26 % a principios del nuevo milenio; ese era el caso particular de Brasil, cuya alta desigualdad está bien documentada, seguido del 20 % en Chile. En ese mismo período, la región menos desigual en el mundo era Europa con cerca de 10 %, mientras que en el Medio Oriente era alrededor de 20 % como tope y en África también, pero con mayor variabilidad. Entonces, para ese periodo, podemos afirmar que un 25 % de participación es lo máximo que el Perú podría tener razonablemente, y en ese caso tendría una desigualdad parecida a la de Brasil. En el período de 2016-2018, la WID.World reporta un modesto incremento en esta misma

participación en Brasil que tiene cerca de 28 %, Chile que ha pasado de 20 % a 24 %⁴ en los Estados Unidos que ha pasado a 20 % y varios países en el Medio Oriente que han pasado la valla del 20 % (Qatar tiene más de 30 %). En África, los países miembros de la OCDE y Asia se mantiene relativamente estable (Rusia disminuye considerablemente esa participación, por ejemplo). La cifra del 25 % podría aumentarse hasta 30 % para este período como máximo para el Perú, pero es un extremo pues lo haría comparable en desigualdad a Qatar. Por completitud, en la Figura A1 en Anexos presentamos la participación del ingreso del 10 % más rico en esos mismos países, para los mismos períodos. Esa figura muestra que, como máximo, en los países con mayor desigualdad (Brasil, países del Medio Oriente y África) se aglomera cerca entre el 50 % y el 70 % entre el 10 % de la población con mayores ingresos, y la misma cifra está entre 30 % y 50 % para los países menos desiguales (principalmente los europeos).

Debemos mencionar que la definición de ingresos detrás de las participaciones de WID.world opera a nivel de individuos y antes de impuestos, así que no son exactamente la misma a nuestra definición de ingreso que es a nivel de hogares y luego de impuestos. No obstante, eso no es un problema para nuestras conclusiones, porque si fuesen comparables y WID.World usara las mismas definiciones que nosotros, las participaciones de WID.World serían menores⁵ y, consecuentemente, los límites creíbles para el Perú que impondríamos para calificar a una probabilidad como no creíble serían aun más bajos que 25 % o 30 %⁶. Así pues,

⁴ Estos dos países son los únicos en América Latina que han recibido el tratamiento DINA en los últimos años; para Argentina y Colombia no se tienen correcciones recientes.

⁵ Esto se debe a que los impuestos tienen un efecto proequidad, en particular en el Perú donde hay alta informalidad y los impuestos son pagados con gente con mayores ingresos. Además, la desigualdad entre hogares (per cápita) es menor que la desigualdad entre individuos, porque el ingreso de un individuo se promedia con el de sus familiares.

⁶ Es decir, eso incrementaría el rango de supuestos que llevan a probabilidades que calificamos como no creíbles y debilita la posición de quienes consideran que el Gini del Perú no ha caído en el tiempo o que es mayor al reportado en la Enaho en grandes magnitudes.

diremos que si una probabilidad requiere que imponamos una participación del 1 % más rico en el Perú que sea mayor a 25 %, no sería muy creíble en vista de esta comparación internacional.

Capítulo III. Discusión Metodológica

1. Modelo de la Distribución de Ingresos

En nuestro modelo subyacente de ingresos asumimos que una fracción (fijada por el investigador) de hogares con los ingresos más altos de la muestra se distribuye de acuerdo con alguna distribución, y que esta no se traslapa con la distribución que caracteriza al resto de la población (i.e. se trata de una partición de la población en dos grupos según niveles de ingreso). Este es el marco fundamental de Atkinson (2007). Bajo estos supuestos, Alvaredo (2011) encuentra que el coeficiente de Gini de la población se puede descomponer, en calidad de identidad, como una suma lineal de tres componentes. Específicamente, es una suma ponderada del Gini de ambos grupos, más un término desigualdad entre ambos grupos. Esta descomposición es una identidad exacta, y su fórmula es

$$G = G(p) = ps G_R + (1 - p)(1 - s)G_N + (s - p) \dots (1)$$

El parámetro p corresponde a la fracción de los hogares más ricos de la población (por ejemplo, si fijamos esta fracción en 2 %, estamos exigiendo que el ingreso de los hogares 2 % más rico de la distribución empírica proviene una Pareto). El número s es la participación en el ingreso total del p % de hogares con mayores ingresos, y el término G_R corresponde al coeficiente de Gini de este grupo. Luego, por construcción, la proporción de hogares restantes sería $(1 - p)$ (i.e. la fracción $(1 - p)$ % con menores ingresos de la población), su participación en el ingreso total sería $(1 - s)$; su coeficiente de Gini es denotado por G_N . El término $(s - p)$ se interpreta como la desigualdad entre las dos particiones de la población.

Todos los elementos hasta ahora mencionados se pueden estimar sin problemas por métodos no paramétricos estándares una vez que uno fija la clasificación de hogares ricos implicada por p . En particular, una manera útil de calcular s es usando la identidad $s = \frac{p\mu_R}{\mu^*}$,

donde μ_R es la media muestral de los ingresos del p % de hogares con mayores ingresos, y $\mu^* = p\mu_R + (1 - p)\mu_N$ es la media de la muestra total. Aquí μ_N es la media muestral de los ingresos del $(1 - p)$ % de hogares con menores ingresos.

Para descomponer el coeficiente de Gini, en principio basta con calcular G_N y G_R de la muestra con cualquier *software* estadístico, pero es una práctica muy común y útil para el análisis asumir que la función de distribución de los hogares con mayores ingresos es conocida. Si esta distribución es de Pareto con parámetros (α, y_m) , se cumple analíticamente que $G_R = 1/(2\alpha - 1)$ y encima $\mu_R = \frac{\alpha y_m}{\alpha - 1}$. En efecto, Atkinson (2007) y Jenkins (2017) comentan que la aplicación predominante consiste en recurrir a la Pareto para describir la cola superior de ingresos. Detrás de esta práctica, sí existen estudios teóricos (Reed, 2003; Charpentier y Flachaire, 2019) y estudios empíricos en países desarrollados (Clementi y Gallegati, 2005) que respaldan la idea de que existe un umbral a partir del cual la distribución de Pareto es una forma efectiva de modelar la distribución de la cola superior de ingresos. No obstante, una vez que adoptamos esta forma funcional para la distribución de la cola superior de ingresos, debemos proponer un método de estimación del parámetro α , y esta decisión no está libre de complicaciones.

Para formalizar la notación requerida para formular estimadores de α , denotamos a la muestra ordenada de ingresos de un par año por $\mathbf{Y} = \{Y_{(1)}, Y_{(2)}, \dots, Y_{(N)}\}$ tal que $Y_{(i)} > Y_{(i+1)}$ para todo $1 \leq i < N$. Al subíndice también se le conoce como rango de la observación. Ahora, dado un valor de la fracción de hogares ricos p , el rango de la observación con ingreso mínimo en la cola superior sería $n_p = pN$. Luego, la submuestra $\{Y_{(1)}, Y_{(2)}, \dots, Y_{(n_p)}\}$ se distribuye de acuerdo con la función de densidad de Pareto:

$$f(Y_{(i)}; \alpha, Y_{(n_p+1)}) = \frac{\alpha Y_{(n_p)}^\alpha}{Y_{(i)}^{\alpha+1}}, \forall Y_{(i)} > Y_{(n_p+1)}$$

El parámetro $Y_{(n_p+1)}$ es un resultado de la partición de la muestra total y, por tanto, viene implicado por la decisión del investigador. El parámetro de forma α , también conocido como índice de cola, puede interpretarse como un indicador que es pequeño si: (i) las colas de la distribución son anchas, (ii) la frecuencia de *outliers* es alta y (iii) la desigualdad es alta (esto se puede ver directamente de la fórmula). Su estimador de máxima verosimilitud (MV) de la distribución de Pareto, conocido como el estimador de Hill, es una de las opciones más usadas en la literatura que usa microdatos (Ibragimov y Ibragimov, 2018) pues su forma funcional y su varianza asintótica son conocidas (Hill, 1975):

$$\alpha_{Hill} = \frac{n_p}{\sum_{i=1}^n [\log Y_{(i)} - \log Y_{(n_p+1)}]}$$

$$s.e.(\alpha_{Hill}) = \alpha_{Hill} / \sqrt{n_p}$$

Recientemente, una rama de la literatura ha criticado la aplicación de la Pareto. Iniciada por Jenkins (2017) y formalizada con simulaciones de Charpentier y Flachaire (2019), los autores argumentan que existe un sesgo por especificación que se ve afectado por las dos decisiones del investigador en su modelo que son materia de su juicio: (I) fracción usada para clasificar hogares como ricos y (II) la distribución empleada para modelar la cola superior de ingresos. En el caso particular en que la distribución usada es la Pareto, Charpentier y Flachaire (2019) y Jenkins (2017) explican que existe un *trade-off* en la elección de la fracción: mientras mayor sea la fracción p de hogares en la cola superior de ingresos, menor será el estimador α_{Hill} . A través de simulaciones de Montecarlo, Charpentier y Flachaire (2019) verifican que en el límite cuando $p \rightarrow 0$, este sesgo se borra, y aunque esto parezca deseable, en la práctica ese límite es muy pequeño y fuera del estándar usado en la literatura ($p \in [0.01, 0.1]$)⁷. A esto se

⁷ Pocos trabajos manejan $p = 0.1$ (Ruiz y Woloszko, 2016), o menos que $p = 0.01$, y usualmente se evalúan valores desde 0.01 hasta 0.05.

suma que un menor p reduce el tamaño de muestra usado para calcular α_{Hill} y aumenta considerablemente la varianza asintótica, y se tiene un *trade-off* entre sesgo y varianza. Además, Charpentier y Flachaire (2019) y Jenkins (2017) encuentran que solo cuando la fracción de hogares ricos p es muy pequeña, los datos de la cola de distribución se asemejan razonablemente bien a la Pareto. En cambio, encuentran que estimaciones de la cola superior usando distribuciones del tipo-Pareto, como la Pareto generalizada y la Pareto extendida, se desempeñan mucho mejor en este rubro, y también son menos sensibles a la elección de p ⁸.

No ignoramos estas advertencias, pero mantenemos la postura de utilizar la distribución de Pareto. Para las correcciones, proponemos dos formas de aliviar este sesgo. Por un lado, al igual que Jenkins (2017), buscaremos verificar si nuestros resultados se mantienen ante diferentes valores de la fracción p ⁹. Por otro lado, usaremos un estimador de α con propiedades que minimizan el sesgo de especificación. Se sabe que el estimador de Hill también puede ser obtenido como resultado de una regresión mínimos cuadrados ordinarios (MCO) entre el logaritmo del rango y el logaritmo del tamaño del ingreso del grupo con mayores ingresos. Gabaix e Ibragimov (2011) demuestran que el componente de primer orden del sesgo se puede eliminar tan solo restando $1/2$ al rango de las observaciones antes de aplicar la regresión. Es decir, nuestro estimador preferido $\hat{\alpha} = \alpha_{Rank-1/2}$ es el estimador MCO de α en la regresión:

$$\log(i - 1/2) = c - \alpha \log(Y_{(i)}) \dots (2),$$

⁸ Esos hallazgos no sorprenden pues están arraigados en el segundo teorema de la teoría de valores extremos, o teorema de Balkema-de Hann (Balkema y de Haan, 1974). Este indica que en toda distribución con colas anchas (incluyendo toda distribución de ingresos), la cola extrema derecha de la distribución converge a alguna distribución de Pareto generalizada. La Pareto extendida, a su vez, es una generalización de la Pareto generalizada, así que el resultado también le aplica.

⁹ Probaremos valores enteros dentro del rango común de análisis en la literatura: $\{0.01, 0.02, \dots, 0.1\}$

usando la submuestra $\{Y_{(1)}, Y_{(2)}, \dots, Y_{(n_p)}\}$. Más aun encuentran una expresión asintótica para el error estándar: $s.e.(\hat{\alpha}) = s.e.(\alpha_{Rank-1/2}) = \hat{\alpha} \sqrt{\frac{2}{n_p}}$. Una aplicación práctica de esta metodología se encuentra en Ibragimov e Ibragimov (2018), para el caso de Rusia.

Otras razones nos inducen a adoptar la distribución de Pareto. Al tratarse de la distribución más estimada en la literatura, aumentamos la comparabilidad de nuestros resultados con el grueso de la literatura. Además, gracias a la simplicidad de la distribución, es posible encontrar una relación analítica directa entre el índice de cola α , y la participación en el ingreso s . Nótese que la identidad de la participación en el ingreso de los hogares con mayores ingresos se puede reescribir como:

$$s = \left[1 + \frac{1-p}{p} \frac{\mu_N}{\mu_R} \right]^{-1}$$

Además, el supuesto de la distribución de Pareto, $\mu_R = \mu_R(\alpha) = Y_{(n_p)} \left[1 - \frac{1}{\alpha} \right]^{-1}$ donde la derivada satisface $\frac{d\mu_R}{d\alpha} < 0$. Luego, es inmediato ver que la participación es una función explícita e invertible del índice de cola: $s = s(\mu_R(\alpha)) = h(\alpha)$. Por regla de la cadena, se tiene que $\frac{dh}{d\alpha} = \frac{ds}{d\mu_R} \frac{d\mu_R}{d\alpha} < 0$. El valor escogido de p no altera el signo de esta relación: mayor desigualdad de ingresos entre los hogares ricos (menor α) conlleva a que estos hogares manejen una mayor proporción del ingreso total. Inclusive sabemos que en el extremo cuando α se acerca a 1 por la derecha¹⁰, la media μ_A se vuelve arbitrariamente grande, y la participación s tiende a 1.

Estas propiedades analíticas nos dicen que, dado una proporción fija p y una muestra de datos y , existe una relación negativa 1 a 1 entre el índice de cola α y la participación en el

¹⁰ En general, el índice de cola $\alpha > 1$, porque de otra forma, la media de la distribución de Pareto no está bien definida.

ingreso s del p % de hogares más ricos. Estas relaciones describen un mapeo invertible y muy sencillo de los supuestos sobre la forma de la distribución hacia la participación en el ingreso s . Por ejemplo, si el investigador considera que la muestra y se distribuye de acuerdo con nuestro modelo con parámetros (p, α) , el valor implicado de la participación de los hogares más ricos que debe aceptar es $s = h(\alpha)$ ¹¹.

Finalmente, el coeficiente de Gini puede agregarse a este conjunto de relaciones cuando la proporción p es particularmente pequeña. De la descomposición original, es inmediato notar que

$$\lim_{p \rightarrow 0} G(p) = (1 - s)G_N + s = f(s)$$

Este supuesto se aplica a nuestros ejercicios, pues nuestros resultados de mayor interés competen a cuando $p \leq 0.05$. Como $f'(s) = 1 - G_N > 0$, notamos que la participación en el ingreso de los hogares ricos es el determinante fundamental de la desigualdad en nuestros ejercicios (G_N ya es observado en la muestra con representatividad, así que está fijo). Asimismo, dada la relación $s = h(\alpha)$, cualquier supuesto sobre α también afecta el valor del coeficiente de Gini, al punto que se puede determinar $\frac{df}{d\alpha} = \frac{df}{dh} \frac{dh}{d\alpha} > 0$ y en el extremo, $\lim_{\alpha \rightarrow 1} f(h(\alpha)) = 1$, que es el caso de desigualdad total.

2. Corrección Semiparamétrica del Coeficiente de Gini

El algoritmo que describimos a continuación sigue muy de cerca la propuesta de Cowell y Flachaire (2007). Utilizamos el estimador $\hat{\alpha}$ de Gabaix e Ibragimov (2011) para especificar a una distribución de Pareto con la cual generamos una nueva muestra de la cola superior de

¹¹ De forma equivalente, si el investigador usa el valor p para particionar la muestra y , y además considera que la que el p % de hogares con mayores ingresos recibe una proporción s del total de ingresos, se puede resolver el valor implicado para α que debe aceptar el investigador con la inversa $\alpha = h^{-1}(s)$.

ingresos, libre de problemas de dispersión y de *outliers*. Además, tomamos una muestra con reemplazo de los ingresos del resto hogares. Repetimos este proceso varias veces, calculando el Gini de la unión de estas muestras de ricos y no ricos en cada iteración, efectivamente implementado un *bootstrap* semiparamétrico que nos permite calcular errores estándares. Esta es la forma correcta de calcular estos errores estándares: ejercicios de simulación muestran categóricamente que el uso de un *bootstrap* estándar resultan en errores estándares severamente subestimados (Cowell y Flachaire, 2007; Davidson y Flachaire, 2007). El algoritmo específicamente es el siguiente:

- 1) Fijamos un número de repeticiones para el “bootstrap”, M , y la fracción de hogares más ricos que se distribuyen Pareto, $p \in (0,1)$.
- 2) Particionamos a la muestra ordenada en submuestras $\{Y_{(1)}, Y_{(2)}, \dots, Y_{(n_p)}\}$ y $\{Y_{(n_p+1)}, Y_{(n_p+2)}, \dots, Y_{(N)}\}$. Con la primera submuestra, estimamos el índice de cola denotado por α_p^* con la regresión MCO de Gabaix e Ibragimov (2011).
- 3) Para cada $l = 1, \dots, M$:
 - a. Generamos N números aleatorios de la uniforme estándar $\{u_1, u_2, \dots, u_N\}$ Ahora:
 - (I) Si $u_i \in [0, 1 - p]$, la i -ésima observación de la muestra es la observación de rango $N(1 - u_i)$; es decir, $Y_i^{(l)} = Y_{(N(1-u_i))}$.
 - (II) Si $u_i \in [1 - p, 1]$, la i -ésima observación de la muestra es tomada aleatoriamente de la distribución Pareto con parámetro α_p^* y observación mínima $Y_{(n_p+1)}$; es decir, $Y_i^{(l)} \sim \text{Pareto}(\alpha_p^*, Y_{(n_p+1)})$.

b. Calcule el coeficiente de Gini $G_p^{(l)}$ de esta muestra $\mathbf{Y}^{(l)}$ de ingresos¹².

4) El Gini corregido será el promedio $G_p^{(0)} = \frac{1}{m} \sum_{l=1}^m G_p^{(l)}$ y su error estándar será

$$s. e. \left(G_p^{(0)} \right) = \frac{1}{m} \sum_{l=1}^m \left(G_p^{(l)} \right)^2 - \left(\frac{1}{m} \sum_{l=1}^m G_p^{(l)} \right)^2.$$

Al final, iteramos pasos 1 al 4 para todos los valores de p deseados, en todos los años de muestra disponibles, y obtenemos series corregidas para diversas combinaciones de supuestos.

3. Algoritmo de Simulación de Montecarlo del Coeficiente de Gini

Adaptamos el algoritmo propuesto por Cowell y Flachaire (2007) a un ejercicio de simulación de Montecarlo. En principio, todos los pasos son fundamentalmente los mismos; la diferencia principal radica en que no estimamos ningún valor para α (bajo el problema del truncamiento, cualquier estimación de este parámetro solo con datos de encuestas está sesgado hacia arriba), sino que lo fijamos en algún valor antes de empezar la generación de números. Además, en las simulaciones generamos simultáneamente muestras *bootstrap* para dos vectores de datos, y contrastamos sus estadísticos simulados en cada iteración. Los vectores de datos pueden ser ingresos de años distintos, pero con el mismo α , o dos réplicas del mismo vector de ingresos, pero en uno elegimos α y en el otro usamos $\hat{\alpha}$ estimado de la muestra. El algoritmo es el siguiente:

1) Fijamos un número de repeticiones para el *bootstrap*, M , la fracción de hogares más ricos que se distribuyen Pareto, $p \in (0,1)$, un índice de cola α_1 para el vector de datos

$$\mathbf{Y}_1 = \{Y_{1(1)}, Y_{1(2)}, \dots, Y_{1(N_1)}\} \quad \text{y otro índice } \alpha_2 \quad \text{para el vector } \mathbf{Y}_2 = \{Y_{2(1)}, Y_{2(2)}, \dots, Y_{2(N_2)}\}.$$

¹² Estrictamente, no es necesario recurrir a la fórmula para hallar un Gini corregido, solo es necesario calcular el Gini sobre la muestra corregida con el método semi-paramétrico. Esto se puede hacer con la fórmula o con un paquete no paramétrica que calcule Gini.

2) Para cada $l = 1, \dots, M$:

a. Generamos N_1 ¹³ números aleatorios de la uniforme estándar $\{u_1, u_2, \dots, u_{N_1}\}$. Ahora para cada uno de los vectores de datos ($j = 1, 2$):

(I) Si $u_i \in [0, 1 - p]$, la i -ésima observación de la l -ésima muestra *bootstrap* $\mathbf{Y}_j^{(l)}$ es la

observación de rango $N_j(1 - u_i)$ de \mathbf{Y}_j ; es decir, $Y_{j,i}^{(l)} = Y_{j,(N_j(1-u_i))}$.

(II) Si $u_i \in [1 - p, 1]$, la i -ésima observación de la muestra es tomada aleatoriamente de la distribución Pareto con parámetro α_j y observación mínima $Y_{j,(n_p+1)}$; es decir,

$$Y_{j,i}^{(l)} \sim \text{Pareto}(\alpha_j, Y_{j,(n_p+1)}).$$

b. Calculamos los coeficientes de Gini de las muestras simuladas $\mathbf{Y}_1^{(l)}, \mathbf{Y}_2^{(l)}$, los denotaremos por $G_1^{(l)}, G_2^{(l)}$.

Acabadas estas simulaciones, se puede calcular una gran variedad de estadísticos y probabilidades para efectuar pseudopruebas de hipótesis. Además de calcular las típicas media y desviación estándar *bootstrap* de cada vector de estadísticos, nos interesa particularmente calcular la probabilidad frecuencial de que el Gini simulado del vector \mathbf{Y}_1 sea mayor que el Gini simulado del vector \mathbf{Y}_2 . En la práctica, esto sería la proporción de casos que este evento sucede en las muestras *bootstrap*:

$$\text{Prob}(\mathbf{G}_1 - \mathbf{G}_2 > c) = \frac{1}{M} \sum_{l=1}^M 1(G_1^{(l)} - G_2^{(l)} > c)$$

Donde $1(\cdot)$ es la función índice. En el caso cuando $\mathbf{Y}_1 = \mathbf{Y}_2$ pero $\alpha_1 \neq \alpha_2 = \hat{\alpha}$ esta sería la probabilidad de que el Gini verdadero (G_1) en un año dado sea mayor al Gini observado en la muestra (G_2) en al menos una magnitud de $c \in [0, 1]$. En el caso cuando $\mathbf{Y}_1 \neq \mathbf{Y}_2$ son vectores

¹³ Sin pérdida de generalidad se tiene que $N_1 \leq N_2$ porque el tamaño de la muestra del año 1 es menor que el tamaño de la muestra del año 2, o bien es la misma cuando son del mismo año.

de ingresos de dos años distintos y $\alpha_1 = \alpha_2 = \alpha$ esta sería la probabilidad de que el Gini en el año del vector 1 sea mayor al Gini del año 2 por al menos c . En la práctica estándar de estadística inferencial, estos eventos deberían ser aceptados si estas probabilidades son por lo menos 0.9. Así pues, nos interesa conocer bajo qué valores de (p, α) , y en particular del segundo, se cumple con este requisito y cuándo no. Dada la relación $s = h(\alpha)$, imponer (p, α) es equivalente a imponer $(p, h(\alpha))$, y para fines de presentación, los resultados serán expuestos en estos términos de $(p, h(\alpha))$.

4. Limitaciones

La principal limitación que enfrenta nuestro trabajo es la incapacidad de ofrecer una corrección que trate el problema de cobertura truncada. Empíricamente solo podemos atender el problema de cobertura dispersa. Enfatizamos que esas correcciones deben interpretarse como cotas inferiores de la desigualdad verdadera, en ese sentido no se oponen a otras correcciones en la literatura peruana. Optamos por las simulaciones para tratar la cobertura truncada, tanto por la falta de datos de impuestos como para dar una perspectiva de análisis novedosa a esta literatura. Pero queda pendiente realizar un esfuerzo por recoger datos de impuestos en el Perú, aunque sean tabulados, o en el mejor de los casos, recolectar datos de impuestos de toda la región de América Latina, y armar estimadores regionales.

Intentamos atender el sesgo por especificación dentro de la posibilidades que ofrece un modelo donde la cola superior es de Pareto. No obstante, siguiendo las ideas de Charpentier y Flachaire (2019) y Jenkins (2017), sería interesante realizar correcciones y simulaciones donde se impone otras distribuciones a la cola superior de ingresos. Creemos que es perfectamente posible adaptar nuestros algoritmos para permitir que la distribución de la cola superior sea una

Pareto generalizada y una Pareto extendida, y sería motivo para un estudio más técnico de la desigualdad.

Finalmente, no estamos implementando ningún peso estadístico en ningún cálculo o estimación que se vea más adelante. Ciertamente, trabajos como los de Alfons, Templ y Filzmoser (2013) o Charpentier y Flachaire (2019) argumentan que es importante incorporar los pesos de las encuestas de hogares en nuestras correcciones. No obstante, dudamos de la utilidad de los pesos de encuestas en una situación donde se reconoce abiertamente que las encuestas exhiben errores sistemáticos en el muestreo. En ese sentido, estamos más alineados con las ideas de Atkinson (2007) y Atkinson *et al.* (2011). En otras palabras: si los pesos fuesen correctos, no habría ningún problema con los estadísticos de las encuestas de hogares luego de incorporarlos, pero el supuesto de partida es que las encuestas de hogares tienen problemas. Dicho de diferente forma: ¿Cómo se puede confiar en que los pesos ofrecen una forma correcta de acercar la distribución muestreada a la verdadera si, por supuesto de partida, creemos que la encuesta de hogar no muestrea (y jamás ha podido muestrear) representativamente a la verdadera distribución de ingresos? Es por consistencia lógica con nuestros supuestos que decidimos no usar pesos¹⁴.

¹⁴ Ya se debe haber hecho evidente que el tema del uso de los pesos genera un debate encendido en la literatura. Autores que crean en la efectividad completa de los pesos también creerán que toda esta literatura es largamente trivial. Por el contrario, los autores que empezaron esta literatura creen fundamentalmente que hay errores en las encuestas de hogares que los pesos simplemente no pueden resolver. Evidentemente, habrá autores en un punto medio entre ambos extremos.

Capítulo IV. Datos y Estadísticos

1. Descripción de las Bases de Datos

En este capítulo presentamos la aplicación de nuestra metodología para el caso peruano. Usamos microdatos a nivel de hogares de la Enaho desde el 1997 hasta 2018, que es producida anualmente por Instituto Nacional de Estadística e Información (INEI). Esta encuesta ha sido la fuente principal para el estudio de los ingresos y gastos de la población peruana por su accesibilidad a todo el público, rango y detalle de preguntas, y gran tamaño de muestra. Es la fuente oficial para la producción de estadísticos del Estado sobre la distribución de ingresos, incluido el Gini. Específicamente usamos el ingreso neto anual del hogar deflactado (`inghog2d`) y el número de miembros del hogar (`mieperho`). Ambos están disponibles en el módulo Sumaria de la Enaho.

También codificamos otras variables de ingreso. Analizamos el ingreso neto anual total del hogar, (`inghog2d` sin ninguna modificación) y el ingreso en escala equivalente (simplificado $\text{inghog2d}/(\text{mieperho}^{0.5})$) para controlar por las economías de escala del ingreso en el número de miembros del hogar. También simularemos coeficientes y tendencias del ingreso bruto anual per cápita del hogar (`inghog1d`), y describiremos las fuertes similitudes entre la tendencia del ingreso neto anual y el ingreso neto laboral anual. Para codificar el ingreso laboral, recogemos los datos individuales de ingreso en el módulo 500 de la Enaho, cuya fuente corresponde al trabajo dependiente e independiente, tanto en la ocupación primaria como secundaria, y hallamos la suma de ingresos laborales de todos los miembros del hogar.

Debemos mencionar que en el 2004 se introdujo un cambio metodológico en la recolección de datos de ingresos y gastos. Por ello, estrictamente los datos antes del 2004 no son completamente comparables con los datos luego del año 2004. Incorporamos los primeros años

de la Enaho igualmente para tratar de visualizar, en alguna medida cualitativa, la tendencia histórica de la desigualdad hasta donde haya datos disponibles, pero en todo momento interpretamos los resultados de cada tramo por separado.

2. Estadísticos Descriptivos de la Desigualdad en el Perú

Como mencionan Ferreira *et al.* (2013) y Genoni y Salazar (2015), dentro de la región de América Latina, el desarrollo económico en el Perú ha sido particularmente excepcional, en tanto este país ha presentado mejoras en varios indicadores claves de desarrollo económico.

Winkelried y Torres (2019), con datos de la Enaho, que en el período entre el 2004 y 2018 el Perú experimentó un ciclo de bonanza económica acompañado de una fuerte reducción en pobreza y un incremento generalizado en movilidad social. A estos estadísticos, uno puede añadir fácilmente y con los mismos datos que la desigualdad medida por el Gini ha caído significativamente de lo que era en el período 2004-2008 a los niveles en que se encuentra en 2018. Esta información se encuentra en el panel (a) de la Figura 2 y en la Tabla 1. En ambos, además, se verifica que la diferencia es estadísticamente significativa. Los motivos de esta caída en desigualdad pueden ser varios, y merecen un estudio aparte específico para el Perú. Sin embargo, una explicación estadística sensata de este resultado la ofrece el panel (b): la participación de una misma fracción (el 1 % o el 2 %) de los hogares más ricos en el ingreso total del país ha disminuido a lo largo de los 15 años de muestra. Manteniendo todo lo demás constante, esto implicaría por la fórmula en (1) que el coeficiente de Gini total debería caer. En la Tabla 1 se puede ver que la participación en el ingreso de los hogares con mayores ingresos cae para diferentes fracciones de hogares que usemos para definir a estos hogares¹⁵.

¹⁵ En cálculos aparte, calculamos el error estándar de todas estas participaciones por el mismo *bootstrap* usado en el Panel (b) y podemos asegurar que la diferencia entre la participación en el 2004 y la participación en el 2018 es estadísticamente significativa.

Tabla 1.*Estadísticos descriptivos de la desigualdad de ingresos en el Perú: 2004-2018*

Año	N	Índice de Gini	Intervalo de confianza – 95 %		Part. Ingreso 1 % más rico	Part. Ingreso 2 % más rico	Part. Ingreso 5 % más rico	Part. Ingreso 10 % más rico
2004	19502	0.509	0.500	0.517	0.114	0.167	0.272	0.391
2005	19895	0.520	0.510	0.530	0.119	0.173	0.281	0.401
2006	20577	0.520	0.512	0.527	0.116	0.170	0.278	0.399
2007	22204	0.527	0.518	0.535	0.121	0.175	0.283	0.404
2008	21502	0.513	0.506	0.521	0.112	0.165	0.272	0.392
2009	21753	0.509	0.501	0.517	0.115	0.168	0.273	0.391
2010	21496	0.488	0.481	0.494	0.105	0.155	0.257	0.372
2011	24809	0.486	0.480	0.493	0.109	0.159	0.259	0.372
2012	25091	0.476	0.469	0.482	0.094	0.143	0.244	0.359
2013	30453	0.475	0.470	0.481	0.092	0.141	0.242	0.359
2014	30848	0.477	0.471	0.483	0.096	0.145	0.247	0.363
2015	32188	0.473	0.468	0.478	0.091	0.140	0.242	0.359
2016	35785	0.470	0.465	0.476	0.089	0.137	0.237	0.355
2017	34584	0.469	0.464	0.474	0.084	0.132	0.234	0.352
2018	37462	0.457	0.453	0.462	0.079	0.125	0.225	0.343

Nota: La variable analizada es Ingreso Anual per cápita de los hogares. El Gini es calculado no paramétricamente

con la función `gini.m` de Matlab ©, y la participación, con la identidad mencionada en la sección III.A. El error

estándar del Gini fue calculado con un “bootstrap” estándar. Fueron empleados todos los hogares, rurales y

urbanos, para el Gini. En cálculos aparte, calculamos el error estándar de todas estas participaciones por *bootstrap*

estándar y podemos asegurar que la diferencia entre la participación en el 2004 y la participación en el 2018 es

estadísticamente significativa.

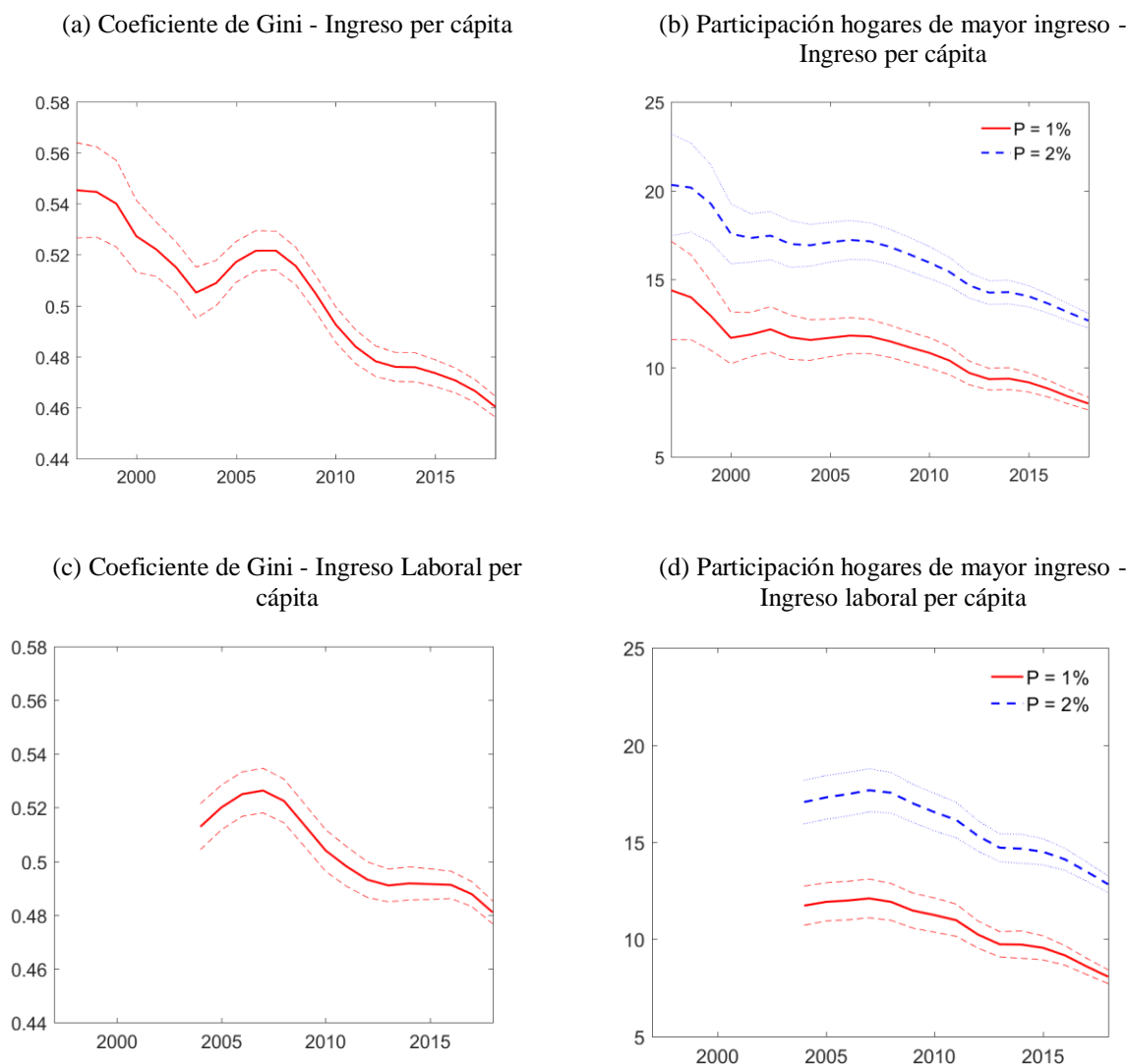


Figura 2. Evolución de estadísticos de desigualdad de ingresos en el Perú: 1997 a 2018. Adaptado de Enaho 1997-2018. La variable analizada es el ingreso neto anual del hogar (inghog2d), dividido entre el número de miembros del hogar (mieperho) en los paneles (a) y (b), y el ingreso laboral anual per cápita en los paneles (c) y (d). El Gini es calculado no paramétricamente. La participación en el ingreso del p % de hogares con mayores ingresos también es calculada no paramétricamente con el procedimiento descrito en la metodología. Los errores estándares para el Gini y la participación de ingresos son calculados por un *bootstrap* estándar con 2,500 simulaciones. Las series han sido suavizadas con un promedio móvil simétrico de tres años.

Sobre el período de 1997 a 2003, el panel (a) sugiere que el Gini ha caído. Pero la caída es modesta y debe ser tomada con pinzas, en vista de la volatilidad de las estimaciones del Gini y el número reducido de hogares en las encuestas previas al año 2000. Además, el panel (a) sugiere

en su totalidad que los niveles de la desigualdad previa al año 2000 son mayores que cualquier año luego del 2000, pero por el cambio metodológico no podemos asegurar este dato. En todo caso, sí está conforme con los pocos hechos estilizados que se conocen de la desigualdad en el Perú antes del 2000. Se considera que en el siglo XX, la región de América Latina en general pudo haber sido una de las más desiguales en el mundo, donde niveles de Gini parecidos a los de África y el Medio Oriente, en el orden de 60 puntos, eran comunes (Gasparini, Cruces y Tornarolli, 2011; Lustig, Lopez-Calva y Ortiz-Juarez, 2013). Mendoza, Leyva y Flor (2011) ofrecen una recopilación con cierta homogeneidad de la desigualdad de ingresos en el Perú desde el año 1980, y en efecto confirman que el coeficiente de Gini registrado por estudios de ese entonces era del orden de 60 puntos.

En suma, estos hechos estilizados conforman una figura muy positiva para el Perú: Hubo un período de alto crecimiento económico entre 2004 y 2018, como es sabido, y al parecer este crecimiento se distribuyó de una forma proequidad en vez de concentrar en los sectores que ya tenían mayores ingresos en la población. Esta sinergia entre crecimiento económico y desigualdad es una afirmación fuerte, y ciertamente vale la pena someter a la mayor cantidad de pruebas posibles.

El primer problema de representación en la Enaho radica en la sobrerrepresentación de los ingresos laborales y la subrepresentación de ingreso por rentas de capital. En efecto, usualmente se cree que las encuestas de hogares tienden a muestrear a trabajadores asalariados pero muy raramente a empresarios dueños de capital grande e inclusive mediano (Székely y Hilgert, 2007). Alarco, Castillo y Leiva (2019) estudian detalladamente la composición del ingreso de los hogares más ricos en la Enaho, y encuentran muchos hechos que se oponen al conocimiento común de lo que debería ser el ingreso de los peruanos más ricos: muy baja

participación del ingreso de capital, relativamente pocos hogares limeños y hogares no muy ricos en general, etc. Para respaldar esta idea, en los paneles (c) y (d) de la Figura 2, calculamos rápidamente la serie de tiempo del coeficiente de Gini de ingreso laboral per cápita y las participaciones del 1 % y 2 % de hogares con mayores ingresos laborales. Inmediatamente se puede visualizar la similitud de la serie del Gini con ingresos totales y con ingresos laborales, no solo en tendencia, sino también en niveles. Esta evidencia sugiere que, en efecto, las encuestas sobremuestrean hogares con trabajadores asalariados, y no capturan bien a los empresarios más adinerados.

3. Evidencia de Correcciones al Coeficiente de Gini en el Perú

Alarco, Castillo y Leiva (2019) proveen un resumen actualizado de todos los esfuerzos realizados para corregir el coeficiente de Gini en el Perú. Todas las correcciones al momento se han basado en aplicar métodos de corrección por cuentas nacionales. En general, se encuentra de forma unánime que el coeficiente de Gini corregido es mayor que aquel reportado por el INEI, pero las magnitudes e interpretaciones del desarrollo del Gini han sido muy diversas. Por un lado, los ajustes de Yamada y Castro (2006), Yamada, Castro y Bacigalupo (2012) y Yamada, Castro y Oviedo (2016) reemplazan el ingreso medio de la Enaho con el de las cuentas nacionales, asumiendo una distribución log-normal al ingreso para redistribuir los excedentes. Encuentran correcciones año a año de altos órdenes de magnitud, pero decrecientes en el tiempo: pasan de aumentar el Gini en 15 puntos porcentuales en 2004 hasta aumentos de 5 puntos porcentuales en 2014. Mas aun, mantienen que la reducción del Gini desde el 2004 ha sido rápida y significativa, incluso luego de las correcciones. Yamada, Castro y Oviedo (2016) enfatizan el rol de las transferencias públicas en esta reducción. Por su lado, Cruz-Saco, Seminario y Campos (2018) y Alarco *et al.* (2019) modelan la cola superior como una Pareto.

Las correcciones de estos dos trabajos son aun más grandes: a lo largo de toda la serie, la corrección está 20 puntos porcentuales (o más) por encima del coeficiente oficial y encima encuentran que la caída en los últimos años no ha sido tan pronunciada o no ha habido.

¿Cómo se puede explicar esta volatilidad en correcciones? Por un lado, ya mostramos que en la literatura internacional se sospecha que las correcciones basadas en cuentas nacionales, incluso las más sofisticadas, pueden sobreestimar el verdadero coeficiente de Gini (Deaton, 2005). Por otro lado, consideramos que esta volatilidad es una manifestación clara del problema de sensibilidad de las correcciones los supuestos metodológicos que necesariamente se deben imponer en el proceso de corrección. Además de usar distribuciones diferentes para imputar el residual del ingreso a la distribución, los trabajos Cruz-Saco, Seminario y Campos (2018) y Alarco *et al.* (2019) imputan nuevos grupos de ingreso a la distribución. Es decir, agregan un nuevo decil de ingreso a la derecha de la distribución (de tal forma que la muestra original ahora corresponde a los nueve primeros deciles), al cual se le imputa el 95 % o el 100 % del residuo entre la media del ingreso en las encuestas y en la cuenta nacional. Un procedimiento de este tipo necesariamente producirá grandes coeficientes de Gini, y motiva el uso de mecanismos para revisar la sensatez de los supuestos de corrección.

Capítulo V. Análisis de las Correcciones

1. Estimaciones del Índice de la Cola

En la Figura 3 presentamos una serie de tiempo del estimador MCO de la regresión rango-tamaño (en rojo), nuestro estimador preferido para el índice de la cola, contrastado con el estimador de Hill (en azul). En la Tabla A1 en Anexos, presentamos las estimaciones en detalle para varios niveles de p , los cuales permiten ver el valor específico de nuestras estimaciones del índice de cola.

En general, el Perú presenta un índice de cola entre 2 y 3. Este valor se encuentra dentro del rango de 1.5 a 3 que, según Ibragimov e Ibragimov (2018), son los valores comunes asociados a países desarrollados. Al analizar las intersecciones de los intervalos de confianza, podemos observar que, para ambos estimadores, el índice de la cola no ha cambiado significativamente excepto los últimos años, donde el índice crece notablemente. Respecto de la interpretación del índice como la inversa del coeficiente de Gini de los hogares con mayores ingresos, esto nos indica que la desigualdad entre estos hogares se ha mantenido relativamente estable a lo largo del período de muestra, incluso es posible que haya caído en los últimos años. Este resultado se mantiene en los cuatro paneles, es decir, aun si cambiamos la clasificación de hogares con mayores ingresos. Con esta evidencia preliminar en consideración, si podemos descartar la idea de que la desigualdad entre los hogares ricos haya aumentado en el período de muestra; nada en la evidencia lo sugiere. La limitación evidente en esta afirmación es que nuestros estimadores están basados únicamente en encuestas de hogares, por lo que pueden estar sobreestimados. En los años previos al 2004, los índices de cola estimados son estadísticamente iguales entre ellos e iguales a los de la mayoría de los años luego del 2004; nuevamente, ello

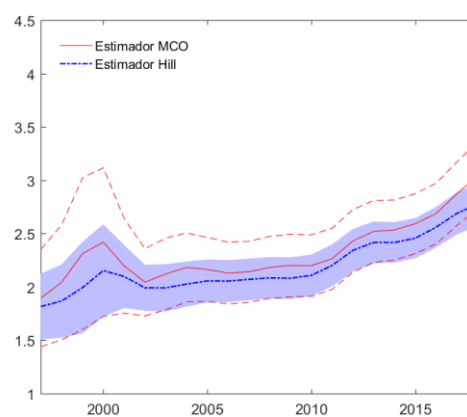
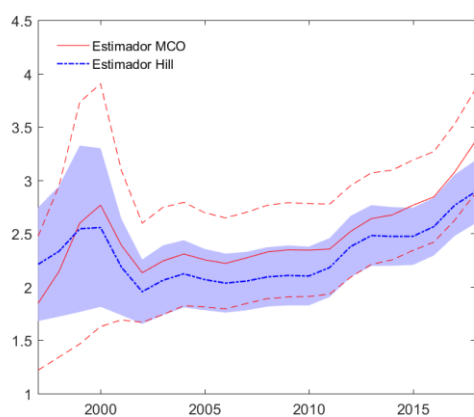
sugiere que la desigualdad entre ricos no ha cambiado. Lamentablemente, los estimadores en esos primeros años exhiben alta volatilidad.

La Figura 3 muestra más resultados de interés técnico. Al comparar los intervalos de confianza azules y rojos dentro de cada panel, encontramos un resultado que esperábamos. Si reducimos el número de hogares con mayores ingresos (como en los paneles (a) y (b) donde solo empleamos a los hogares en el 1 % y 2 % más alto de la distribución de ingresos, respectivamente), encontramos que el estimador de Hill y el estimador MCO son básicamente los mismos y la varianza es particularmente grande. En cambio, cuando el número de hogares con mayores ingresos es relativamente grande (paneles (c) y (d)), la varianza se reduce considerablemente y, además, las diferencias entre MCO y Hill empiezan a ser más grandes. En particular, consistentemente se halla que el estimador puntual MCO es mayor que el de Hill (i.e. la desigualdad entre ricos es menor si es estimada por MCO que si es estimada por Hill). En los últimos años del panel (d), la diferencia ya es estadísticamente significativa. Este fenómeno es una manifestación del *trade-off* entre sesgo y varianza de los estimadores del índice de Pareto, y más aun de que el estimador MCO sí puede ofrecer una corrección de la subestimación del estimador de Hill (Ibragimov e Ibragimov, 2018; Gabaix e Ibragimov, 2011).

Nos interesa saber qué tan confiable son nuestros estimadores, en el sentido de conocer si son sensibles a la clasificación de hogares con mayores ingresos. La Figura 4 muestra una pequeña prueba de robustez de ambos estimadores para las muestras de 4 años particulares. La prueba exige que los intervalos de confianza de un estimador se intercepten a medida que se incrementa la muestra del grupo de hogares con mayores ingresos usados para la estimación, pasando desde el 1 % hasta el 10 % de los hogares más ricos de la muestra. Para las muestras del 2004, 2008 y 2013 se puede ver que esta prueba es satisfecha (todos los intervalos de confianza

graficados se intersecan dentro de cada panel). Esto significa que el estimador del índice es robusto a la decisión de especificación de cuántos y cuáles hogares se considera que se distribuyen Pareto. Solo anotamos que en el 2018 esta prueba no se satisface, pues el índice de cola estimado con el 1 % de hogares más ricos es mayor que el índice de cola estimado con el 10 % de hogares más ricos. Esto también sucede en menor medida, en el 2015 y el 2017.

(a) Índice de cola – 1 % hogares con mayores ingresos (b) Índice de cola – 2 % hogares con mayores ingresos



(c) Índice de cola – 5 % hogares con mayores ingresos

(d) Índice de cola – 10 % hogares con mayores

ingresos

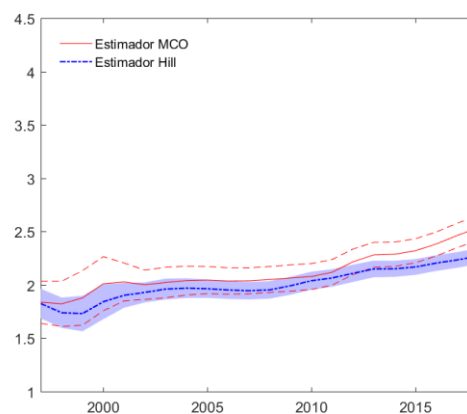
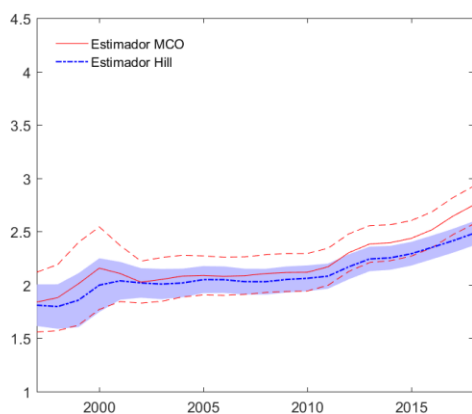


Figura 3. Serie de tiempo de las estimaciones del índice de la cola superior del ingreso para diferentes clasificaciones de hogares con mayores ingresos. Adaptado de Enaho 2004-2018. La variable analizada es el ingreso neto anual del hogar, dividido entre el número de miembros del hogar. Los índices y sus errores estándares son

calculados de acuerdo con las fórmulas en el texto. Fueron empleados todos los hogares, rurales y urbanos, que clasifican como hogares en el grupo con mayores ingresos. Para clasificar a estos hogares, cada panel emplea una diferente fracción de hogares más ricos en la estimación de los índices de cola. El título de cada panel indica cómo se hizo la clasificación: En el panel (a), por ejemplo, se emplearon el 2 % de los hogares con los ingresos más altos de toda la muestra.

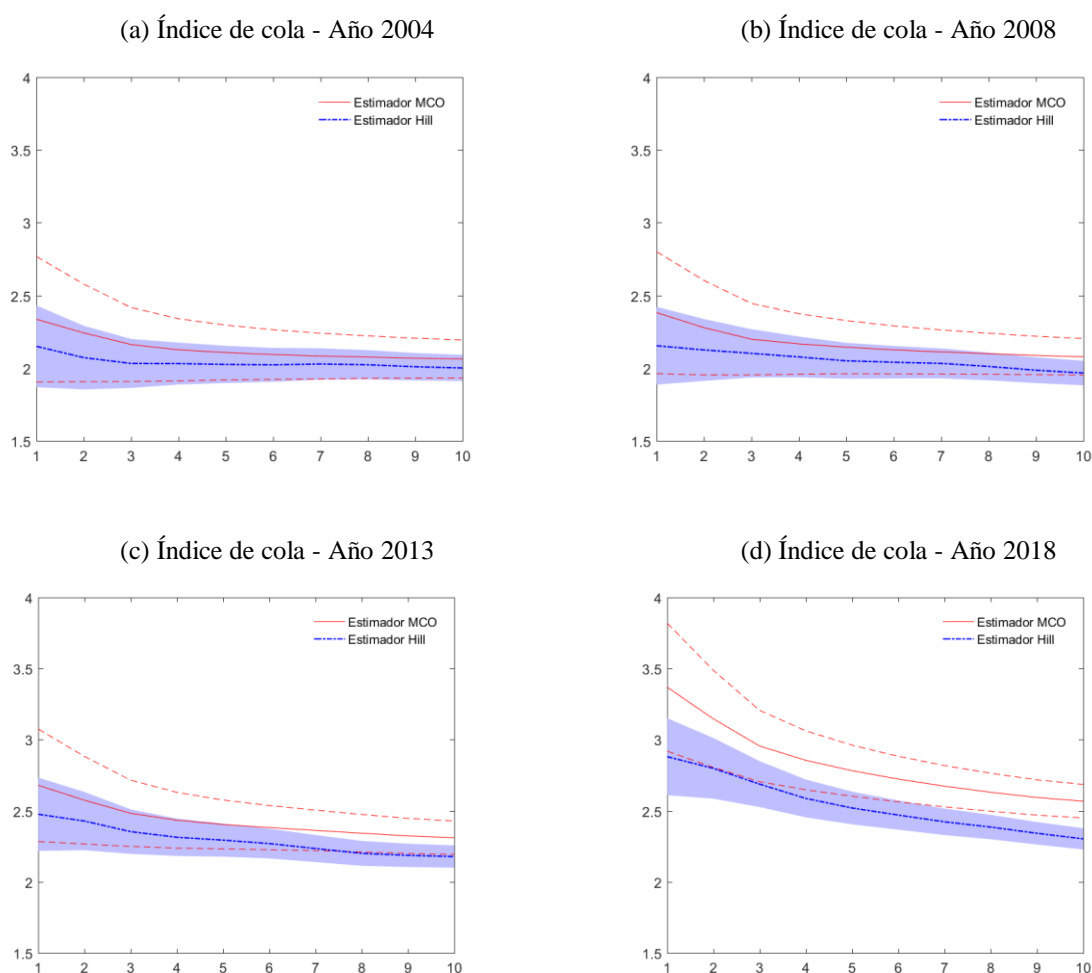


Figura 4. Estimaciones del índice de la cola superior de ingresos para diferentes años de la Enaho y diferentes proporciones de hogares no observados. Adaptado de la Enaho 2004-2018. La variable analizada es el ingreso neto anual del hogar, dividido entre el número de miembros del hogar. Los índices y sus errores estándares son calculados de acuerdo con las fórmulas en el texto. Fueron empleados todos los hogares, rurales y urbanos, que clasifican como hogares con en el grupo con mayores ingresos. Para clasificar a estos hogares, el eje horizontal de cada gráfico indica la fracción de hogares con ingresos más altos de la muestra que fueron usados en la estimación. Para cada panel, se emplea un año (o ronda) diferente, indicado por el título.

2. Estimaciones de las Correcciones por Cobertura Dispersa

En la Figura 5, en el panel (a), presentamos (en azul) la serie de tiempo del coeficiente de Gini corregido año a año por el algoritmo de corrección que presentamos en la sección III.B, considerando que el grupo de mayores ingresos son el 2 % de hogares con mayores ingresos. También presentamos (en rojo) la serie de tiempo del Gini sin corregir; es exactamente la misma serie del panel (a) de la Figura 2. En general, las correcciones no producen cambios significativos en los estadísticos sin correcciones. Es decir, si el problema de representatividad en la Enaho se redujese únicamente a un problema de dispersión entre el 2 % de hogares más ricos de la muestra, los coeficientes de Gini oficiales serían estimadores insesgados del coeficiente de Gini verdadero. Esto corresponde a la interpretación de Lustig (2019) de que la cobertura dispersa afecta únicamente a la volatilidad de los estadísticos.

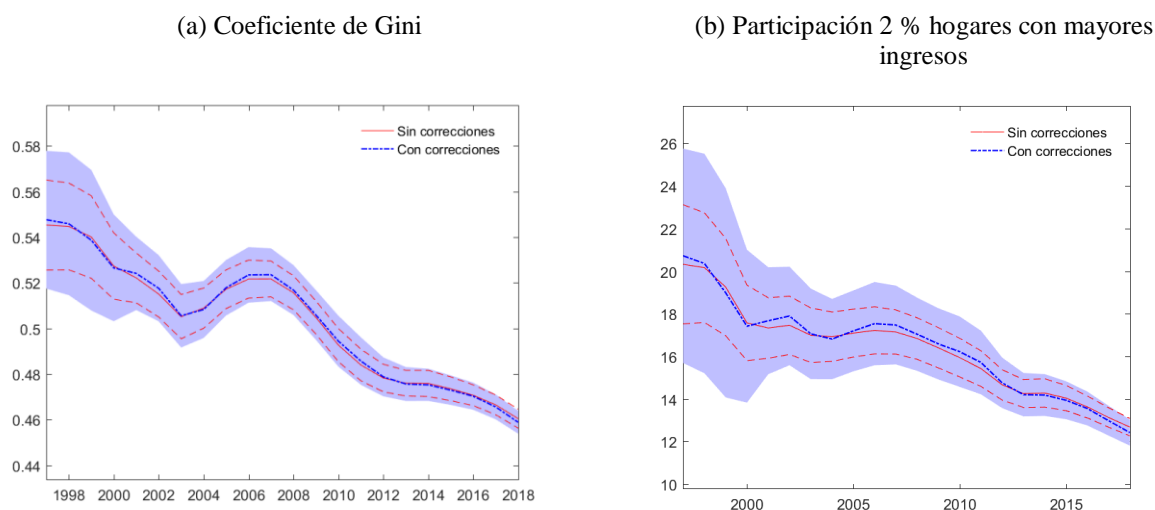


Figura 5. Serie de tiempo de estadísticas de la desigualdad de ingresos, con y sin correcciones. Adaptado de Enaho 1997-2018. La variable analizada es el ingreso neto anual del hogar, dividido entre el número de miembros del hogar. Para clasificar, un hogar está en el grupo de mayores ingresos si se encuentra en el 2 % más rico de la distribución de ingresos. Los estadísticos no paramétricos son exactamente los mismos que los de la Figura 2; la novedad está en la corrección semi-paramétrica que son calculados según lo descrito en el algoritmo de la

metodología. Los errores estándares de los estadísticos no paramétrico son calculado por *bootstrap* estándar; los errores estándares de los estadísticos semiparamétricos, por el *bootstrap* descrito en la metodología.

Para indagar sobre la validez de nuestras correcciones, realizamos pruebas de robustez estándares en la literatura. En la Figura 6, buscamos implementar la misma prueba de robustez que realizamos para los índices de la cola superior de ingresos en la Figura 4. En un solo gráfico, la Figura 6 muestra los estimados corregidos del coeficiente de Gini, cuando se aplican diferentes clasificaciones del grupo de hogares con mayores ingresos (tabuladas en el eje horizontal), y para tres años diferentes de la muestra. Nos interesa que los intervalos de confianza de los estadísticos corregidos para diferentes clasificaciones de hogares ricos se intersecten. En efecto, para cada año, los intervalos se intersectan si se emplea desde el 1 % hasta el 10 % de los hogares más ricos para estimar la cola superior de ingresos.

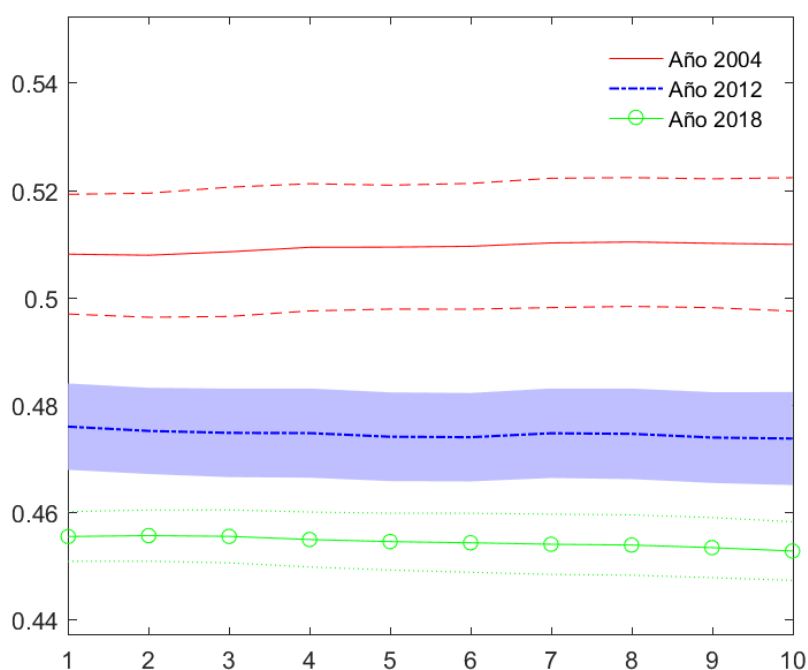


Figura 6. Corte transversal del coeficiente de Gini corregido. Adaptado de Enaho 2004-2018. La variable analizada es el ingreso neto anual del hogar, dividido entre el número de miembros del hogar. El eje horizontal marca la fracción de los hogares más ricos usada para clasificar hogares con mayores ingresos en la estimación

semiparamétrica. El rango de fracciones usadas en el eje horizontal representa las prácticas más comunes en la literatura. Los estimadores puntuales y sus errores estándares son calculados según lo descrito en la metodología, y para cada estimado puntual se le ha agregado un intervalo de confianza al 95 %.

También evaluamos si las correcciones son similares ante cambios sensatos en la variable de análisis. En todo nuestro análisis previo, hemos usado el ingreso neto anual per cápita de los hogares. En las Figuras A2 y A3 de los Anexos, repetimos el mismo ejercicio para ingreso neto anual y el ingreso neto anual equivalente de los hogares, respectivamente. Presentamos la serie de tiempo del coeficiente de Gini corregido, contra el coeficiente sin corregir, en el panel (a), y el corte transversal del coeficiente de Gini corregido (i.e. el coeficiente corregido de un año fijo, pero con diferentes clasificaciones de hogares ricos) del 2004, 2012 y 2018 en el panel (b). Se puede notar que cualitativamente se obtienen los mismos resultados que en la Figura 5 y la Figura 6. Estas pruebas de robustez sugieren que las correcciones son válidas, dada la validez de los supuestos en que están basados (en particular, si el problema en la Enaho es solo de cobertura dispersa y no de truncamiento).

Sin embargo, en vista de las correcciones previamente realizadas en la literatura peruana donde encuentran correcciones de alta magnitud en el coeficiente de Gini, solo podemos afirmar que nuestras correcciones ofrecen una cota inferior razonable del verdadero coeficiente de Gini (es decir, el coeficiente de Gini verdadero necesariamente es al menos tan alto como el oficial). Esto sugiere que la Enaho, así como las encuestas de hogares en todo el mundo, por sí sola no es lo suficientemente representativa para realizar correcciones de los estadísticos de desigualdad sin el apoyo de fuentes externas. Como mencionamos en la sección anterior, esto puede deberse en buena medida a la sobrerrepresentación de los ingresos laborales por encima de las ganancias de capital.

Capítulo VI. Simulaciones del Coeficiente de Gini

Sintetizamos la información de nuestras simulaciones en la forma de probabilidades de ocurrencia de ciertos eventos clave. Estas probabilidades son la proporción de veces que el evento de interés ocurre a lo largo de las M repeticiones de la simulación descrita en el Capítulo III. Si las simulaciones muestran que cierto evento ocurre con al menos 0.9 de probabilidad, podemos afirmar que, bajo los supuestos de esa simulación, ese evento es significativo. Específicamente los supuestos que variamos en cada simulación son la proporción de hogares no observados y el índice de cola de estos hogares no observados, que juntos implican una participación en el ingreso de estos hogares no observados.

1. Simulaciones de las Correcciones Anuales del Coeficiente de Gini

Primero realizamos simulaciones de la probabilidad de observar correcciones positivas y de diversas magnitudes en el coeficiente de Gini. La Figura 7 muestra las probabilidades de observar cualquier corrección positiva (al menos mayor a 0) para varios años seleccionados de la Enaho. En este tipo de figuras, el eje vertical indica la probabilidad simulada de observar que el Gini bajo los supuestos de esa simulación sea mayor que el Gini simulado con los estimadores de la muestra; además, el eje horizontal y el color de cada línea indican la participación en el ingreso de los hogares ricos y la proporción de hogares no observados asumidos para el cálculo de esa probabilidad, respectivamente¹⁶. Las líneas verticales señalan la participación del 1 %, 2 % y 5 % de hogares con mayores ingresos según los datos de la Enaho. La figura muestra que no se requiere asumir grandes variaciones en la participación de ingresos de los hogares con mayores ingresos para obtener correcciones significativas del coeficiente Gini. Para cualquiera

¹⁶ En la programación probamos valores de $p = \{1, 2, 5\}$ y varios valores del índice de cola α desde 1.1 hasta 4; dados estos valores y la muestra, la distribución está completamente especificada y podemos resolver la participación de ingreso del p % de hogares con mayores ingresos s

de los tres valores propuestos de la proporción no observada de hogares, incrementos pequeños en la participación de ingresos respecto a su valor en la muestra original elevan la probabilidad de obtener una corrección positiva hasta 1 en todos los años, con particular fuerza en el año 2018.

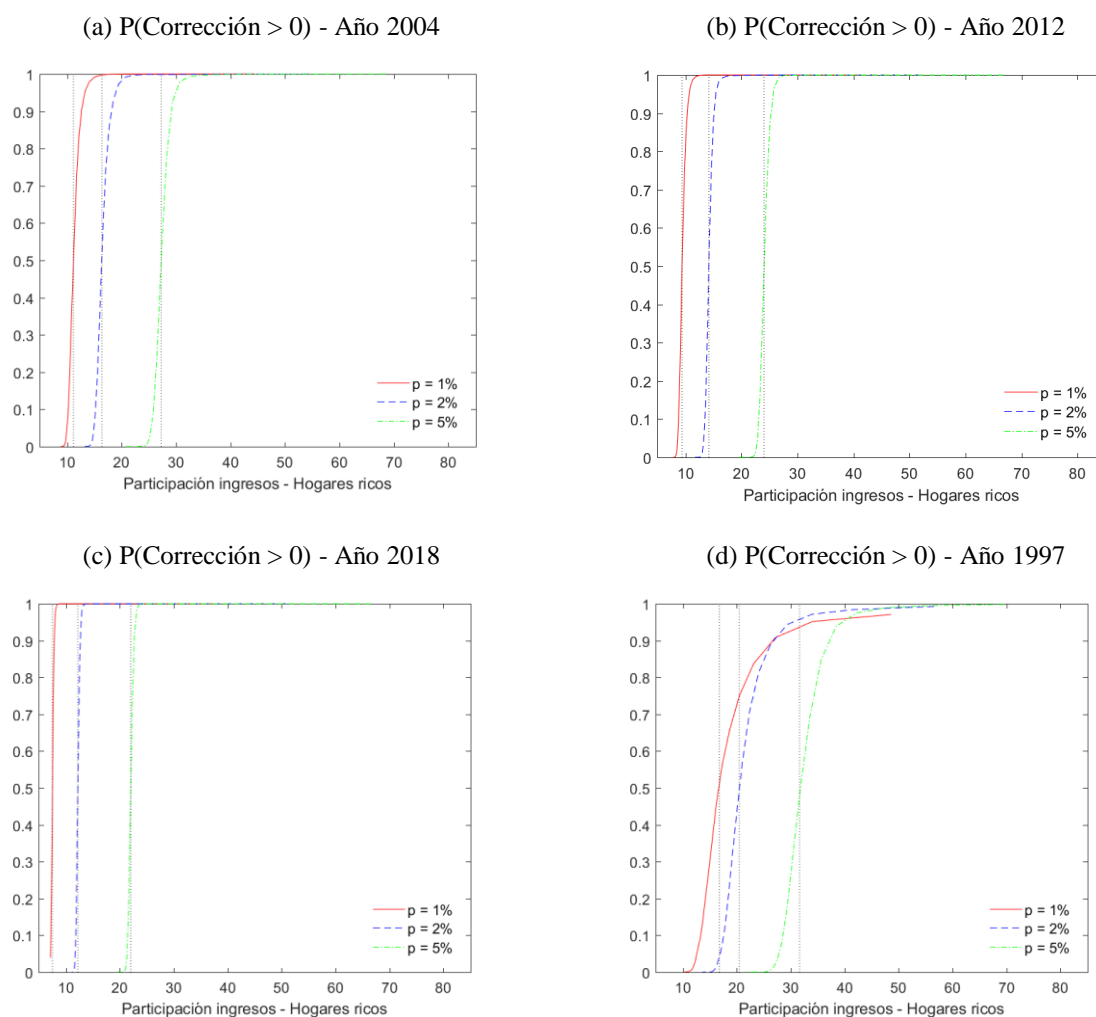


Figura 7. Probabilidades simuladas de observar una corrección positiva en el coeficiente de Gini. Adaptado de Enaho 1997-2018. La variable analizada es el ingreso neto anual del hogar, dividido entre el número de miembros del hogar. El eje vertical indica la probabilidad simulada de observar que el Gini bajo los supuestos de esa simulación sea mayor que el Gini simulado con los estimadores de la muestra. La probabilidad es la proporción de veces que el evento de interés (observar un Gini hipotético mayor al de la muestra) ocurre en 1000 repeticiones del algoritmo descrito en el Capítulo III. El eje horizontal y el color de cada línea indican la participación en el ingreso

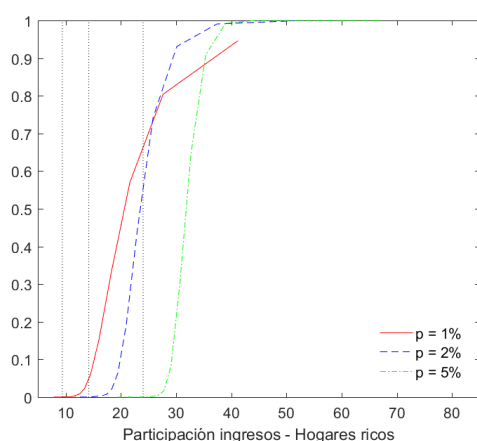
de los hogares ricos y la proporción de hogares no observados asumidos para el cálculo de esa probabilidad, respectivamente. En la programación probamos valores de $p = \{1,2,5\}$ y varios valores del índice de cola α desde 1.1 hasta 4; dados estos valores y la muestra, la distribución está completamente especificada y podemos resolver la participación de ingreso del p % de hogares con mayores ingresos s . Las líneas verticales punteadas señalan la participación del 1 %, 2 % y 5 % de hogares con mayores ingresos según los datos de la Enaho como referencia.

No requerimos recurrir a la valla del 25 % para determinar que es plausible admitir que el verdadero coeficiente de Gini en el Perú es, en alguna magnitud, mayor al reportado en la Enaho, pero sí nos será útil para ver qué tanto mayor puede ser esta corrección. Para ello, en la Figura 8 reportamos la probabilidad de que el coeficiente de Gini simulado sea mayor en 0.05, 0.1, 0.15 y 0.2 al de la muestra para el año 2012. En el panel (a), vemos que, si asumimos que el 1 % de hogares más ricos no es observado, para obtener una probabilidad de 0.9 de que el coeficiente de Gini simulado sea mayor que el de la muestra en al menos 0.05, uno debe creer que ese 1 % de hogares maneja cerca del 40 % de ingresos, que no consideramos creíble. La probabilidad se modera mucho más si se asume que el 2 % o es el 5 % no se observa. En esos casos, la proporción de ingresos de los ricos requerida para obtener una probabilidad de 0.9 no está tan lejos de su estimador.

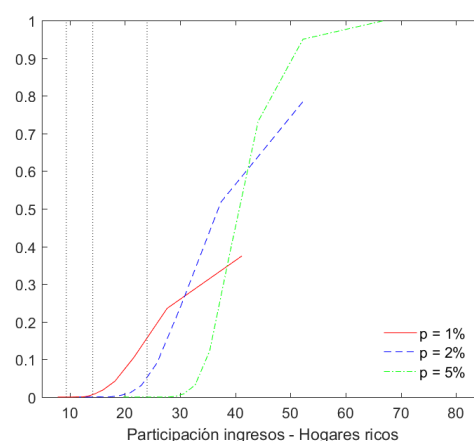
En cambio, en el panel (b), no es posible justificar una probabilidad de 0.9 de obtener una corrección de 0.1 en el coeficiente de Gini: Cuando el 1 % y 2 % de hogares no son observados, ninguna proporción de ingresos que hemos considerado conlleva a esa probabilidad. Aun con 5 % de hogares no observados, para obtener esa probabilidad de 0.9 tendríamos que creer que ese 5 % de hogares no observado controla más del 50 % de ingresos. Para los paneles (c) y (d), en ningún escenario simulado podemos aceptar la hipótesis de que el verdadero coeficiente de Gini sea mayor en 0.15 y 0.2, respectivamente, al coeficiente oficial de la Enaho.

Por eso consideramos que la diferencia entre el verdadero Gini y el de la Enaho en el 2012 debe ser a lo más de 0.1, lo cual es menor que las correcciones encontradas en otros trabajos de la literatura peruana, donde abundan las correcciones de órdenes mayores a 0.1. Este resultado no es el caso aislado del 2012. En las Figura A4, A5 y A6 de los Anexos presentamos las mismas simulaciones que la Figura 8, pero para los años de 2004, 2018 y 1997, respectivamente. Repitiendo el mismo análisis podemos llegar a la misma conclusión de que las correcciones como máximo serían de 0.1 en esos años también (un poco menos en 1997, un poco más en 2018). Como ya hemos mencionado, consideramos que la alta magnitud en las correcciones por cuentas nacionales se debe, al menos en parte, a un caso de sobreestimación de la corrección. Estas simulaciones sugieren exactamente ello, y que las correcciones deberían interpretarse cuidadosamente como cotas superiores de una verdadera corrección; coincidimos plenamente en que sí existe y positiva.

(a) $P(\text{Corrección} > 0.05)$ - Año 2012



(b) $P(\text{Corrección} > 0.10)$ - Año 2012



(c) $P(\text{Corrección} > 0.15)$ - Año 2012

(d) $P(\text{Corrección} > 0.20)$ - Año 2012

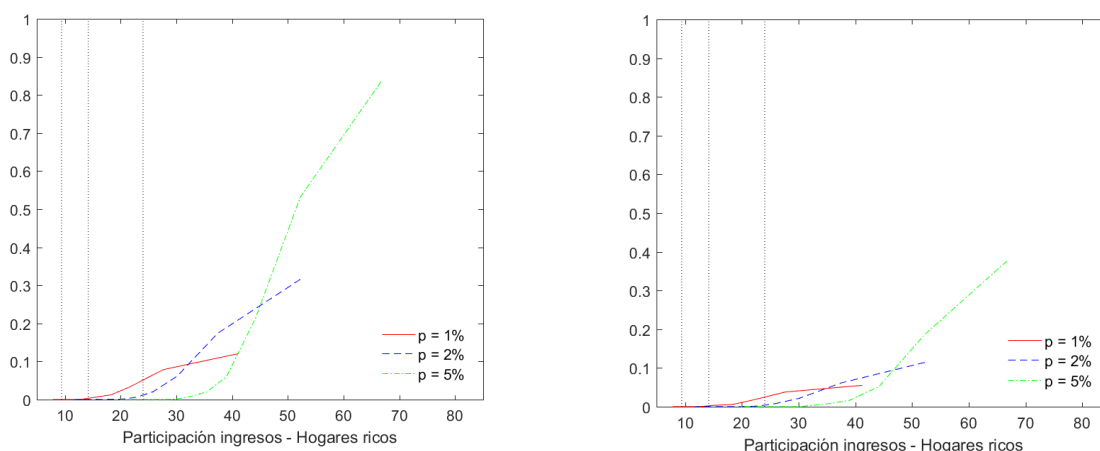


Figura 8. Probabilidades simuladas de observar correcciones de diversa magnitud en el coeficiente de Gini del año 2012. Adaptado de Enaho 2012. La variable analizada es el ingreso neto anual del hogar, dividido entre el número de miembros del hogar. El eje vertical indica la probabilidad simulada de observar que el Gini bajo los supuestos de esa simulación sea mayor que el Gini simulado con los estimadores de la muestra. La probabilidad es la proporción de veces que el evento de interés (observar un Gini hipotético mayor al de la muestra) ocurre en 1000 repeticiones del algoritmo descrito en el Capítulo III. El eje horizontal y el color de cada línea indican la participación en el ingreso de los hogares ricos y la proporción de hogares no observados asumidos para el cálculo de esa probabilidad, respectivamente. En la programación probamos valores de $p = \{1,2,5\}$ y varios valores del índice de cola α desde 1.1 hasta 4; dados estos valores y la muestra, la distribución está completamente especificada y podemos resolver la participación de ingreso del p % de hogares con mayores ingresos s . Las líneas verticales punteadas señalan la participación del 1 %, 2 % y 5 % de hogares con mayores ingresos según los datos de la Enaho como referencia.

2. Simulaciones del Cambio del Coeficiente de Gini en el Tiempo

Utilizamos la misma técnica de simulación para evaluar la probabilidad de observar una disminución en el coeficiente de Gini en dos momentos distintos. La descripción del algoritmo de simulación para este ejercicio se encuentra en el Capítulo III, pero es importante recalcar que, para simular los coeficientes en dos momentos distintos, especificamos el mismo índice de cola a la distribución de cada año. Además, esto ayuda a simplificar la dimensionalidad de nuestro análisis y presentar gráficamente los resultados, la razón principal para hacer esto está en la Figura 3, que claramente muestra evidencia, aunque sea solo de la Enaho, que el índice de cola

no ha cambiado en el período de análisis. Podría haber aumentado en los últimos años, pero nada sugiere que la desigualdad entre los hogares con mayores ingresos haya cambiado.

En la Figura 9 presentamos nuestras simulaciones principales sobre la probabilidad de observar una reducción del coeficiente de Gini entre 2004 y 2018. En la línea roja, se asume que no se observa al 1 % de hogares con mayores ingresos y el eje horizontal es la participación en el ingreso en el año 2004 de ese 1 % asumida para esa simulación (en la línea azul, no se observa el 5 % y el eje horizontal es la participación en 2004 de ese 5 %). Notamos que para obtener una probabilidad menor a 0.9 de observar el evento descrito cuando no se observa el 1 %, se debe asumir que el primer percentil de la distribución controla por lo menos el 25 % de los ingresos total. Esto es justamente el número que pondría al Perú en el nivel de Brasil, los países del Medio Oriente y los africanos en términos de desigualdad, como se vio en la Figura 1. Análogamente para obtener una probabilidad menor a 0.9 de observar el evento descrito cuando no se observa el 5 %, se debe asumir que este 5 % de la distribución controla, por lo menos, una proporción cercana al 40 % de los ingresos totales. Como en los países más desiguales el 10 % maneja del 50 % al 60 % de esas cifras, no es difícil imaginar que estos supuestos también nos pondrían en esa categoría. En la medida en que situar al Perú a ese nivel de desigualdad no sea creíble, podemos afirmar que la reducción en el Gini entre 2004 y 2018 es robusta a críticas de subrepresentación.

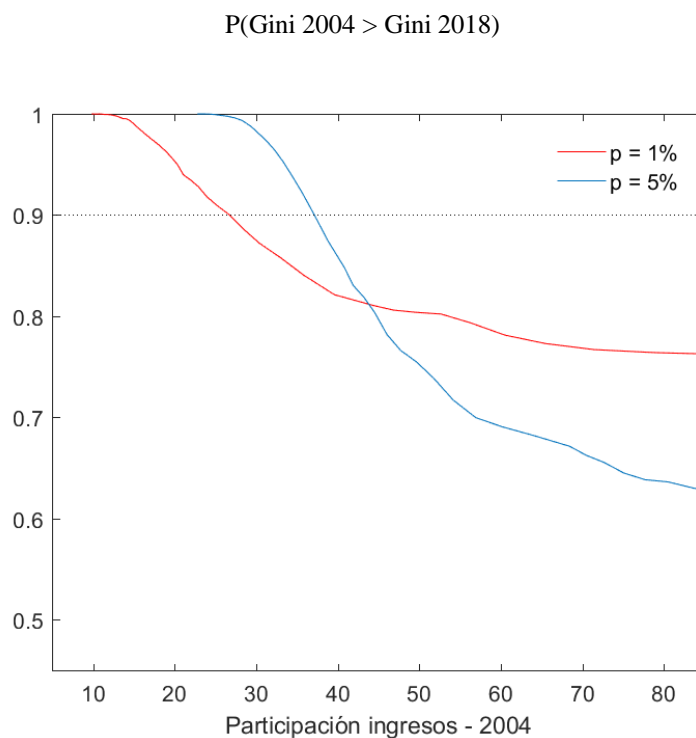
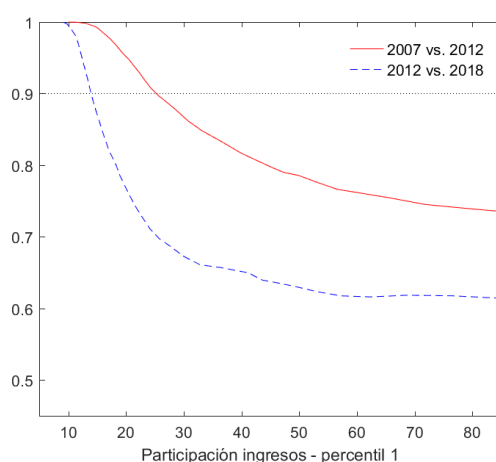


Figura 9. Probabilidades simuladas de observar una reducción del coeficiente de Gini entre 2004 y 2018. Adaptado de Enaho 2004-2018. La variable analizada es el ingreso neto anual del hogar, dividido entre el número de miembros del hogar. El eje vertical indica la probabilidad simulada de observar que el Gini en el 2004 sea mayor que el Gini en el 2018. La probabilidad es la proporción de veces que este evento de interés ocurra en 5,000 repeticiones del algoritmo descrito en el capítulo III. El eje horizontal y el color de cada línea indican la participación en el ingreso de los hogares ricos del 2004 y la proporción de hogares no observados asumidos para el cálculo de esa probabilidad, respectivamente. Si bien no presentamos la participación asumida para el 2018, en todas las simulaciones, fijamos el mismo índice de cola para simular el Gini de cada año, y luego resolvemos el nivel de participación en 2004; de esa forma también se podría resolver el nivel de participación del 2018. Los valores del índice de cola α que probamos van desde 1.01 hasta 3.

Para analizar con más detalle qué está sucediendo, vemos dos ejercicios adicionales. En primer lugar, en la Figura 10, simulamos la misma probabilidad de observar reducciones, pero en dos subperiodos: de 2007 a 2012, y de 2012 a 2018. Realizamos esto porque desde la Figura 2 se observa que el cambio en el coeficiente de Gini oficial se concentró principalmente entre los años 2007 y 2012 (de hecho, el coeficiente de Gini aumentó de 2004 a 2007), y la tendencia se

aplanó considerablemente entre el 2012 y 2018. La Figura 10 corresponde directamente con este hecho. Las participaciones requeridas para obtener una baja probabilidad (i.e. menor a 0.9) de reducción entre 2007 y 2012 son casi las mismas que las de la Figura 9, y mucho más que las que se requieren para obtener una baja probabilidad de reducción entre 2012 y 2018. Esto sugiere que toda la significancia de la reducción en el coeficiente de Gini entre 2004 y 2018 se concentra en la rápida e importante caída producida entre 2007 y 2012.

(a) $P(\text{Gini Año 1} > \text{Gini Año 2})$, $p=1$ % de hogares no son observados



(b) $P(\text{Gini Año 1} > \text{Gini Año 2})$, $p=5$ % de hogares no son observados

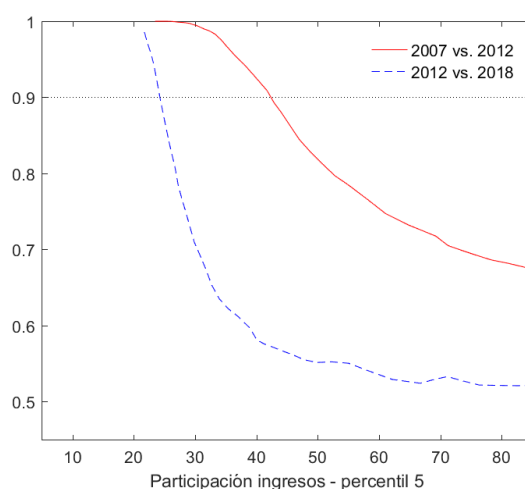


Figura 10. Probabilidades simuladas de observar una reducción del coeficiente de Gini por subperíodos. Adaptado de Enaho 2007-2018. La variable analizada es el ingreso neto anual del hogar, dividido entre el número de

miembros del hogar. El eje vertical indica la probabilidad simulada de observar que el Gini en el primer año sea mayor que el Gini en el segundo año. La probabilidad es la proporción de veces que este evento de interés ocurra en 5000 repeticiones del algoritmo descrito en el Capítulo III. El eje horizontal y el color de cada línea indican la participación en el ingreso de los hogares ricos del primer año y la proporción de hogares no observados asumidos para el cálculo de esa probabilidad, respectivamente. Si bien no presentamos la participación asumida para el segundo año, en todas las simulaciones, fijamos el mismo índice de cola para simular el Gini de cada año, y luego resolvemos el nivel de participación en el primer año; de esa forma también se podría resolver el nivel de participación del segundo año. Los valores del índice de cola α que probamos van desde 1.01 hasta 3.

En segundo lugar, nos gustaría visualizar de qué forma operan nuestros supuestos para afectar la trayectoria de nuestras simulaciones. Para ello, la Figura 11 presenta trayectorias simuladas de los coeficientes de Gini. Las trayectorias graficadas recogen el promedio de las simulaciones del coeficiente Gini para todos los años de la muestra que emplean los mismos supuestos comunes. Por ejemplo, en el panel (a), la línea roja es un gráfico de líneas del promedio de los coeficientes de Gini simulados que resultan de asumir el primer percentil de hogares no se observa en la Enaho y que se distribuyen Pareto con índice de cola $\alpha = 1.5$. Los paneles (c) y (d), además, agregan pseudo-intervalos de confianza al 95 %¹⁷ a sus trayectorias para visualizar la volatilidad de las simulaciones. Los paneles muestran que: (i) Mayor p asumido resulta en correcciones de mayor magnitud. (ii) Las trayectorias simuladas bajo cualquier α o p son marcadamente similares en tendencia a la serie oficial de la Figura 2, panel (a), apenas aplanando la trayectoria en los últimos años. (iii) Consecuentemente, disminuciones en α (y, por tanto, aumentos en la participación en el ingreso de los ricos) producen básicamente traslaciones aproximadamente paralelas y hacia arriba de la trayectoria del coeficiente de Gini. Entonces, si las trayectorias solo se mueven de forma paralela, ¿por qué la probabilidad de

¹⁷ Simplemente son los coeficientes promedio sumado (restado) con 1.96 veces la desviación estándar de la simulación en ese año y bajo esos supuestos.

observar un cambio positivo no es más cercana a 1? Los paneles (c) donde $\alpha = 1.5$ y (d), donde $\alpha = 2.5$, dan una respuesta tentativa a esta pregunta, pues demuestran claramente que (iv) la volatilidad de la simulación incrementa fundamentalmente al asumir menores niveles de α . Ya sabíamos que bajos niveles de α inducen volatilidad en el coeficiente de Gini, y eso está jugando un papel importante en estas simulaciones, pues aumenta artificialmente la volatilidad de la diferencia del coeficiente de Gini entre dos años distintos, y disminuye el número de veces que observamos reducciones en las simulaciones, pero en el promedio podemos ver claramente una tendencia decreciente.

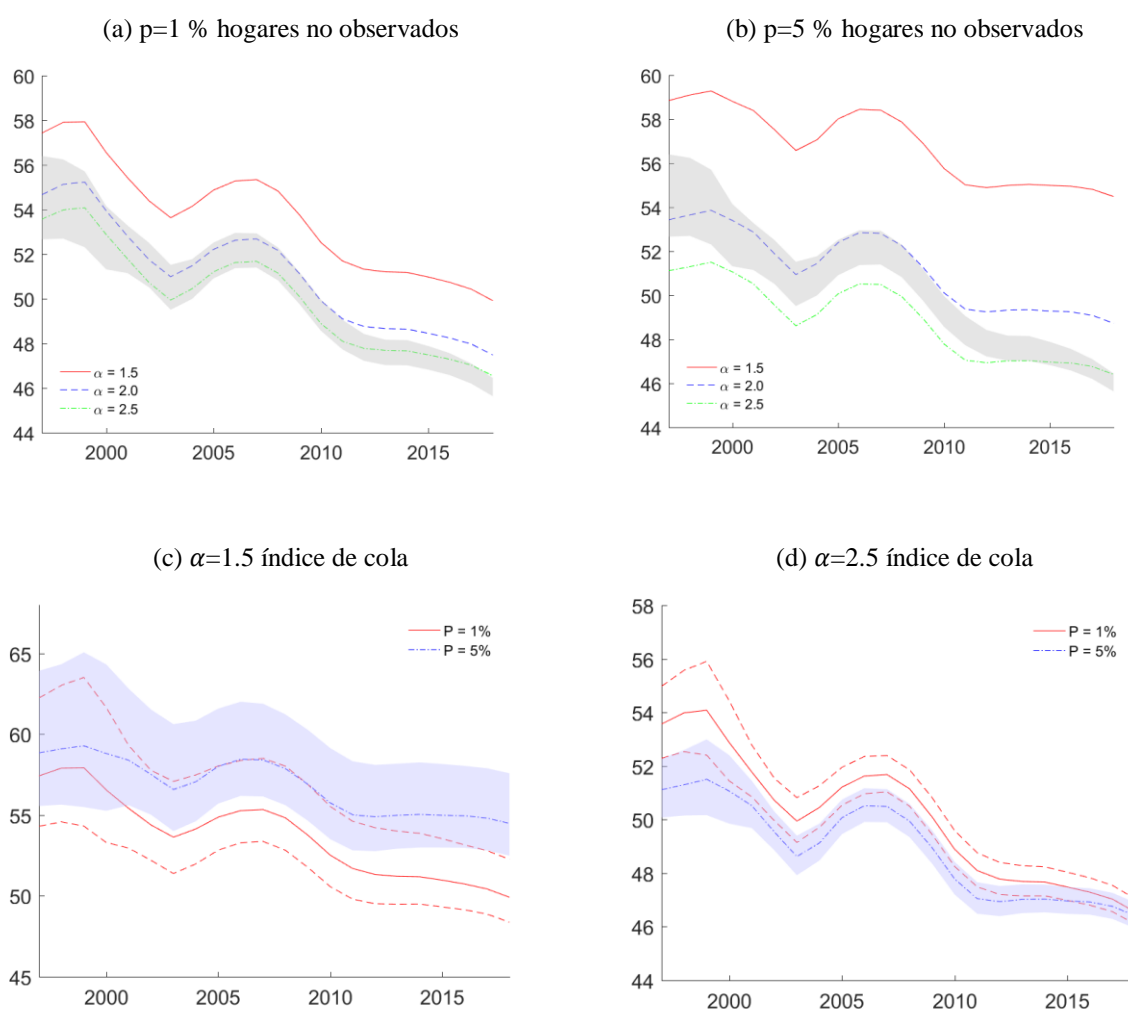


Figura 11. Trayectorias simuladas del coeficiente de Gini bajo supuestos comunes. Adaptado de Enaho 1997-2018.

La variable analizada es el ingreso neto anual del hogar, dividido entre el número de miembros del hogar. El eje

vertical indica el promedio de 5,000 iteraciones del coeficiente Gini simulados bajo los supuestos de α y p indicados en la leyenda y en el título del panel, y para el año indicado en el eje horizontal, usando el algoritmo del Capítulo III. Los paneles (a) y (b) presentan un área gris resaltada que corresponde al intervalo de 95 % de confianza del coeficiente de Gini oficial, tomado de la Figura 2. Los paneles (c) y (d) también incluyen aproximaciones del intervalo de confianza al 95 % del promedio de la simulación, pero principalmente están ahí para visualizar la volatilidad de la simulación. Las series han sido suavizadas con un promedio móvil simétrico de tres años.

Finalmente, en la Figura A7 de los Anexos, presentamos varios de los mismos ejercicios de simulación del cambio del coeficiente de Gini de los ingresos antes de impuestos. Esta definición de ingreso nos interesa porque es más parecida a aquella empleada por la literatura, y es importante saber qué tanto la aplicación de impuestos altera nuestros resultados. Afortunadamente, la Figura A7 es muy similar a las Figuras 9 y 10, así que los supuestos sobre las participaciones asumidas para el ingreso neto aplican a los supuestos que tendrían que asumirse sobre el ingreso antes de impuestos. Además, en el panel (d) de la Figura A7 presentamos las probabilidades de observar una reducción del coeficiente de Gini entre 1997 y 2003. Para obtener una probabilidad de reducción menor a 0.9, se requiere asumir participaciones en el ingreso un tanto menores a 25 %. En todo caso, debemos mencionar que la poca cantidad de datos en 1997 inflan la volatilidad de los coeficientes de Gini simulados.

3. ¿Qué nos Dicen las Simulaciones sobre el Gini del Perú?

De estas simulaciones se pueden rescatar algunas ideas centrales sobre la evolución de la desigualdad en el Perú. Coincidimos con la afirmación de que el coeficiente de Gini en las encuestas de hogares subestima al verdadero coeficiente en alguna medida. Sin embargo, la diferencia entre el Gini verdadero y el oficial no puede ser de un orden mayor a 10 puntos porcentuales. Ningún escenario de simulación nos lleva a aceptar ese evento. Las correcciones en

el resto de América Latina también son de magnitudes pequeñas en comparación a la literatura peruana.

Segundo, sobre el punto que más urge discutir para la política pública: ¿Se puede afirmar que la desigualdad en el Perú cayó luego del 2004? Buena parte de esta sección se ha dedicado a argumentar a favor de esta hipótesis: Los supuestos que se requieren para poder negarla están fuera de lo razonable para el caso peruano, aun siendo un país de una de las regiones con mayor desigualdad en la historia. Consideramos que nuestro enfoque para lidiar con esta pregunta, más allá de vadear el problema de falta de información de impuestos en el Perú, es útil e innovador pues reconoce que no existe una sola corrección correcta y perfecta del coeficiente de Gini, y aborda directamente esta incertidumbre.

Conclusiones y Recomendaciones

Nuestra revisión de literatura de desigualdad de ingresos nos hacía sospechar que se subestimaba, año a año, el coeficiente de Gini en el Perú. Nos propusimos realizar correcciones de fácil aplicación para atender el problema específico de la cobertura dispersa de hogares con mayores ingresos en la Enaho del Perú, entre 1997 y 2018. Los coeficientes corregidos no muestran cambios significativos, y nos limitamos a interpretar por ello que el coeficiente de Gini oficial es una cota inferior del verdadero coeficiente. Dada la evidencia internacional y local, advertimos que la Enaho no ofrece una buena representación de los hogares con mayores ingresos. Seguidamente, propusimos un ejercicio de simulación para pensar sobre la cola superior de ingresos no observada del problema de truncamiento. De acuerdo con estas simulaciones, establecemos que el coeficiente de Gini oficial subestima el verdadero coeficiente en órdenes de magnitud no menores, pero estas correcciones son, como máximo, del orden de 0.1 y, además, la tendencia decreciente en el tiempo es robusta a las críticas por subrepresentación que se le adjudican a la Enaho. Los supuestos requeridos para rechazar este último hecho implicarían que el Perú ostentaría niveles de desigualdad tan altos o hasta mayores que los países reconocidos como los más desiguales en el mundo.

Las conclusiones son relevantes para hacedores de política económica en toda la región y, evidentemente, más para el Perú. Nuestro trabajo está enfocado en ayudar al hacedor de política que, sin acceso a otros datos además de la encuesta de hogares local. El objetivo es saber más sobre la confiabilidad de sus datos sobre la desigualdad, e informar correctamente a la población de hechos estilizados de la desigualdad de ingresos. Nuestras estimaciones sugieren que las encuestas de hogares no terminan de representar bien a los hogares con mayores ingresos. La literatura también informa sobre el rol crítico que juegan los supuestos, inclusive *priors*, de los

investigadores en la implementación y los resultados mismos de sus correcciones, y el cuidado que se debe tener en la presentación de esos resultados.

En la literatura de desigualdad hay muchos campos de estudio prometedores. Un trabajo futuro inmediato implicaría extender este ejercicio a una multitud de países de todo América Latina. Asimismo, tanto para el Perú como en toda la región, queda pendiente hacer un esfuerzo regional por recolectar datos de impuestos, aunque sean tabulados en vez de microdatos, y aplicar la metodología DINA a toda la región, como se ha hecho ya en otras regiones enteras con grandes dificultades para el acceso de datos (Assouad, Chancel y Morgan, 2018; Chancel, Cogneau, Gethin y Myczkowski, 2019; Alvaredo, Assouad y Piketty, 2019). Nuestro método de simulación también goza de amplio espacio para mejorar y ser aplicado. Las grandes virtudes de este método incluyen reconocer que no existe una sola corrección correcta y perfecta del coeficiente de Gini e identificar los parámetros claves que determinan al coeficiente. Debido a que no existe un estimador de desigualdad perfecto ni perfectamente estimado, tal y como Lustig (2019) sugiere, una aproximación prudente que debería seguirse en adelante es la presentación de rangos de desigualdad posibles (o sensatos) en vez de estimadores puntuales.

Referencias

- Alarco, G., Castillo, C. y Leiva, F. (2019). *Riqueza y desigualdad en el Perú: Visión Panorámica*. Lima: Oxfam.
- Alfons, A., Templ, M. y Filzmoser, P. (2013). Robust estimation of economic indicators from survey samples based on Pareto tail modelling. *Journal of the Royal Statistical Society, Series C*, 62(2), 271-286.
- Alvaredo, F. (2011). A note on the relationship between top income shares and the Gini coefficient. *Economics Letters*, 110(3), 274-277.
- Alvaredo, F., & Londoño-Vélez, J. (2013). *High Incomes and Personal Taxation in a Developing Economy: Colombia 1993-2010*. CEQ Working Paper no. 12. New Orleans, LA: CEQ Institute, Tulane University.
- Alvaredo, F., Chancel, L., Piketty, T., Saez, E. y Zucman, G. (2017). Global Inequality Dynamics: New Findings from WID.world. *American Economic Review*, 107(5), 404-409.
- Alvaredo, F., Assouad, L. y Piketty, T. (2019). Measuring Inequality in the Middle East 1990–2016: The World’s Most Unequal Region? *Review of Income and Wealth*, 65(4), 685-711.
- Amarante, V. (2016). Income inequality in Latin America: A factor component analysis. *Review of Income and Wealth*, 62, S4-S21.
- Anand, S. y Segal, P. (2017). Who Are the Global Top 1%? *World Development*, 95, 111-126.
- Assouad, L., Chancel, L. y Morgan, M. (2018). Extreme Inequality: Evidence from Brazil, India, the Middle East, and South Africa. *AEA Papers and Proceedings*, 108, 119-123.

- Atkinson, A. B. (2005). Top Incomes in the UK over the 20th Century. *Journal of the Royal Statistical Society. Series A*, 168(2), 325-343.
- Atkinson, A. B. (2007). Measuring Top Incomes: Methodological Issues. En A. B. Atkinson, & T. Piketty, *Top incomes over the twentieth century: A contrast between continental European and English-speaking countries*, (págs. 18-42). Oxford: Oxford University Press.
- Atkinson, A. B. y Piketty, T. (2010). *Top Incomes. A Global Perspective*. Oxford: Oxford University Press.
- Atkinson, A. B., Piketty, T. y Saez, E. (2011). Top Incomes in the Long Run History. *Journal of Economic Literature*, 49(1), 3-71.
- Bach, S., Corneo, G. y Steiner, V. (2009). From Top to Bottom. The Entire Income Distribution in Germany, 1992-2013. *Review of Income and Wealth*, 55(2), 303-330.
- Balkema, A. A. y de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2(5), 792-804.
- Blanchet, T., Chancel, L. y Gethin, A. (2019). *How Unequal Is Europe? Evidence from Distributional National Accounts, 1980-2017*. World Inequality Lab. WID.world WORKING PAPER N° 2019/06.
- Bourguignon, F. (2015). Appraising income inequality databases in Latin America. *Journal of Economic Inequality*, 13(4), 557-578.
- Burkhauser, R. V., Feng, S., Jenkins, S. P. y Larrimore, J. (2012). Recent Trends in Top Income Shares in the USA: Reconciling Estimates from March CPS and IRS Tax Return Data. *Review of Economics and Statistics*, 94(2), 371-388.

- Burkhauser, R. V., Herault, N., Wilkins, R. K. y Jenkins, S. P. (2016). *What has been happening to UK income inequality since the mid-1990s? Answers from reconciled and combined household survey and tax return data*. ISER Working Paper no. 2016-03. Colchester: Institute for Social and Economic Research.
- [CEPAL] Comisión económica para América Latina y el Caribe. (2019). *Panorama Social de América Latina 2019*. Santiago de Chile.
- Chancel, L., Cogneau, D., Gethin, A. y Myczkowski, A. (2019). *How Large Are African Inequalities? Towards Distributional National Accounts in Africa, 1990-2017*. World Inequality Lab. WID.world WORKING PAPER N° 2019/13.
- Chancel, L. y Piketty, T. (2019). Indian Income Inequality, 1922-2015: From British Raj to Billionaire Raj? *Review of Income and Wealth*, 65, S33-S62.
- Charpentier, A. y Flachaire, E. (2019). *Pareto Models for Top Incomes*.
- Clementi, F. y Gallegati, M. (2005). Pareto's Law of Income Distribution: Evidence for Germany, the United Kingdom, and the United States. En A. Chatterjee, S. Yarlagadda, & B. K. Chakrabarti, *Econophysics of Wealth Distributions* (págs. 3-14). Milano: Springer.
- Cowell, F. A. y Flachaire, E. (2007). Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics*, 141(2), 1044-1072.
- Cruz-Saco, M. A., Seminario, B. y Campos, C. (2018). Desigualdad (Re)considerada Peru 1997-2015. *Journal of Economics, Finance and Internation Business*, 2(1), 13-52.
- Davidson, R. y Flachaire, E. (2007). Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics*, 141(1), 141-166.

- Deaton, A. (2005). Measuring Poverty in a Growing World (or Measuring Growth in a Poor World). *Review of Economic and Statistics*, 87(1), 1-19.
- Díaz-Bazán, T. (2015). *Measuring Inequality from Top to Bottom*. World Bank Policy Research Paper no. 7237. Washington, DC: Banco Mundial.
- Ferreira, F. H., Messian, J., Rigolini, J., López-Calva, L. F., Lugo, M. A. y Vakis, R. (2013). *Economic Mobility and the Rise of the Latin American Middle Class*. Washington D.C.: World Bank.
- Flores, I., Sanhueza, C., Atria, J. y Mayer, R. (2019). Top Incomes in Chile: A Historical Perspective of Income Inequality, 1964-2017. *Review of Income and Wealth*.
- Gabaix, X. y Ibragimov, R. (2011). Rank — $1/2$: A Simple Way to Improve the OLS Estimation of Tail Exponents. *Journal of Business & Economic Statistics*, 29(1), 24-39.
- Gasparini, L., Cruces, G. y Tornarolli, L. (2011). Recent Trends in Income Inequality in Latin America. *Economía*, 11(2), 147-201.
- Genoni, M. y Salazar, M. (2015). Steering toward Shared Prosperity in Perú. En L. Cord, M. Genoni, & C. Rodríguez-Castelán, *Shared prosperity and poverty eradication in Latin America and the Caribbean* (págs. 269-301). Washington, D.C.: Banco Mundial.
- Higgins, S., Lustig, N. y Vigorito, A. (2018). *The rich underreport their income: Assessing bias in inequality estimates and correction methods using linked survey and tax data*. ECINEQ WP 2018-475. ECINEQ.
- Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *Annals of Statistics*, 3(5), 1163-1174.

- Hubert, M., Dierckx, G. y Vanpaemel, D. (2013). Detecting influential data points for the Hill estimator in Pareto-type distributions. *Computational Statistics and Data Analysis*, 65, 13-28.
- Ibragimov, M. y Ibragimov, R. (2018). Heavy tails and upper-tail inequality: The case of Russia. *Empirical Economics*, 54, 823–837.
- Jenkins, S. P. (2017). Pareto Models, Top Incomes and Recent Trends in UK Income Inequality. *Economica*, 84(334), 261-289.
- Kim, N. N. (2018). Top incomes in Korea: Update, 1933-2016. *Korean Economic Forum*, 11(1), 1-21.
- Lakner, C. y Milanovic, B. (2016). Global income distribution. From the fall of the Berlin Wall to the Great Recession. *World Bank Economic Review*, 30, 203-232.
- Lustig, N., Lopez-Calva, L. F. y Ortiz-Juarez, E. (2013). Declining Inequality in Latin America in the 2000s: The Cases of Argentina, Brazil, and Mexico. *World Development*, 44, 129-141.
- Lustig, N. (2019). *The “Missing Rich” in Household Surveys: Causes and Correction Approaches*. CEQ Institute. CEQ Working Paper 75.
- Mendoza, W., Leyva, J. y Flor, J. L. (2011). La distribución del ingreso en el Perú: 1980–2010. *Revista CIS*, 9(15), 27-50.
- Morgan, M. (2018). *Income inequality, growth and elite taxation in Brazil: new evidence combining survey and fiscal data, 2001–2015*. International Policy Centre for Inclusive Growth. Working Paper number 165.

- Novokmet, F., Piketty, T. y Zucman, G. (2018). From Communism to Capitalism: Private versus Public Property and Inequality in China and Russia. *AEA Papers and Proceedings*, 108, 109–113.
- Piketty, T. (2003). Income Inequality in France, 1901–1998. *Journal of Political Economy*, 111(5), 1004-1042.
- Piketty, T. y Saez, E. (2003). Income Inequality in the United States, 1913-1998. *Quarterly Journal of Economics*, 118(1), 1-41.
- Piketty, T., Saez, E. y Zucman, G. (2018). Distributional National Accounts: Methods and Estimates for the United States. *Quarterly Journal of Economics*, 133(2), 553–609.
- Piketty, T., Yang, L. y Zucman, G. (2019). Capital Accumulation, Private Property, and Rising Inequality in China, 1978–2015. *American Economic Review*, 109(7), 2469–2496.
- Reed, W. J. (2003). The Pareto Law of Incomes— An Explanation and an Extension. *Physica A: Statistical Mechanics and its Applications*, 319, 469-486.
- Ruiz, N. y Woloszko, N. (2016). *What do household surveys suggest about the top 1% incomes and inequality in OECD countries?* Economics Department Working Paper no. 1265. Paris: OECD.
- Székely, M. y Hilgert, M. (2007). What's Behind the Inequality We Measure? An Investigation Using Latin American Data. *Oxford Development Studies*, 197-217.
- Szekely, M. y Mendoza, P. (2017). Declining inequality in Latin America: structural shift or temporary phenomenon? *Oxford Development Studies*, 45(2), 204-221.
- The Economist. (26 de Octubre de 2019). Days and Nights of rage in Chile. *The Economist*.
Obtenido de <https://www.economist.com/the-americas/2019/10/26/days-and-nights-of-rage-in-chile>

- Vermuelen, P. (2018). How Fat is the Top Tail of the wealth distribution? *Review of Income and Wealth*, 64(2), 357-387.
- Winkelried, D. y Torres, J. (2019). Economic mobility along the business cycle. The case of Peru. *Applied Economics*. doi:10.1080/00036846.2018.1529401
- Yamada, G., Castro, J. F. y Bacigalupo, J. L. (2012). Desigualdad monetaria en un contexto de rápido crecimiento económico: El caso reciente de Perú. *Revista de Estudios Económicos*, 24, 65-77.
- Yamada, G. y Castro, J. F. (2006). *Poverty, inequality, and social policies in Peru: As poor as it gets*. Centro de Investigación de la Universidad del Pacífico. Lima: Documento de Discusión DD/07/06.
- Yamada, G., Castro, J. F. y Oviedo, N. (2016). *Revisitando el coeficiente de Gini en el Perú: El rol de las políticas públicas en la evolución de la desigualdad*. Centro de Investigación de la Universidad del Pacífico. Lima: Documento de discusión 1606.

Anexos

Tabla A1.

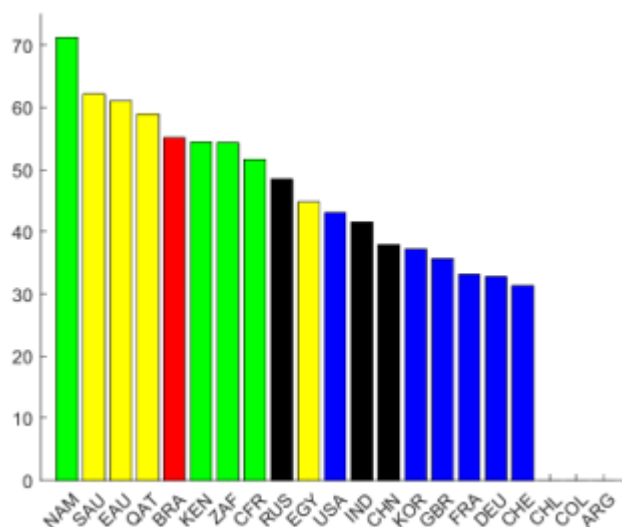
Estimaciones del índice de la cola superior de ingresos en Perú: 2004-2018

Año	% Hogares Ing. Altos	Nº Hogares Ing. Altos	Estimador Rango - Tamaño	Intervalo de confianza - 95%		Estimador de Hill	Intervalo de confianza - 95%	
2004	1	195	2.374	1.902	2.845	2.188	1.881	2.495
2004	5	975	2.107	1.920	2.294	2.029	1.902	2.156
2004	10	1950	2.061	1.931	2.190	1.998	1.909	2.087
2005	1	199	2.205	1.772	2.639	2.005	1.726	2.283
2005	5	995	2.075	1.892	2.257	2.048	1.921	2.175
2005	10	1990	2.035	1.908	2.161	1.944	1.858	2.029
2006	1	206	2.229	1.798	2.659	2.075	1.791	2.358
2006	5	1029	2.095	1.914	2.277	2.070	1.943	2.196
2006	10	2058	2.048	1.922	2.173	1.968	1.883	2.053
2007	1	222	2.215	1.803	2.627	1.986	1.725	2.247
2007	5	1110	2.052	1.882	2.223	2.004	1.886	2.122
2007	10	2220	2.014	1.896	2.133	1.925	1.845	2.005
2008	1	215	2.426	1.967	2.884	2.168	1.878	2.458
2008	5	1075	2.143	1.962	2.324	2.040	1.918	2.162
2008	10	2150	2.074	1.950	2.198	1.960	1.877	2.042
2009	1	218	2.239	1.818	2.659	2.050	1.777	2.322
2009	5	1088	2.082	1.907	2.257	2.032	1.911	2.153
2009	10	2175	2.036	1.915	2.157	1.964	1.881	2.046
2010	1	215	2.485	2.015	2.955	2.160	1.871	2.449
2010	5	1075	2.154	1.972	2.337	2.095	1.970	2.221
2010	10	2150	2.110	1.984	2.236	2.077	1.989	2.164
2011	1	248	2.169	1.787	2.550	2.034	1.781	2.287
2011	5	1240	2.079	1.915	2.243	2.021	1.909	2.134
2011	10	2481	2.053	1.939	2.168	2.035	1.955	2.115
2012	1	251	2.592	2.139	3.046	2.496	2.187	2.805
2012	5	1255	2.362	2.177	2.547	2.187	2.066	2.308
2012	10	2509	2.250	2.126	2.375	2.111	2.028	2.193
2013	1	305	2.722	2.290	3.154	2.481	2.202	2.759
2013	5	1523	2.402	2.231	2.573	2.282	2.167	2.396
2013	10	3045	2.306	2.190	2.422	2.173	2.095	2.250
2014	1	308	2.523	2.125	2.922	2.465	2.190	2.741
2014	5	1542	2.364	2.197	2.531	2.220	2.109	2.331
2014	10	3085	2.259	2.146	2.372	2.139	2.063	2.214
2015	1	322	2.926	2.474	3.377	2.479	2.208	2.750
2015	5	1609	2.450	2.280	2.619	2.291	2.179	2.403
2015	10	3219	2.320	2.207	2.434	2.150	2.076	2.225
2016	1	358	2.686	2.292	3.079	2.469	2.213	2.725

2016	5	1789	2.479	2.317	2.641	2.363	2.253	2.472
2016	10	3579	2.376	2.266	2.486	2.228	2.155	2.301
2017	1	346	3.077	2.618	3.535	2.840	2.541	3.139
2017	5	1729	2.657	2.480	2.835	2.387	2.274	2.499
2017	10	3458	2.455	2.339	2.571	2.206	2.133	2.280
2018	1	375	3.461	2.965	3.956	2.908	2.614	3.203
2018	5	1873	2.778	2.600	2.955	2.513	2.399	2.627
2018	10	3746	2.559	2.443	2.675	2.290	2.217	2.363

Nota: Adaptado de Enaho 2004-2018. La variable analizada es ingreso anual per cápita de los hogares. Los índices y sus errores estándares son calculados de acuerdo con las fórmulas en el texto. Fueron empleados todos los hogares, rurales y urbanos, que clasifican como hogares con mayores ingresos: Son parte del p % más rico de la población; el p está indicado en la segunda columna.

(a) Promedio de la participación en el ingreso nacional del 10% más rico, 2000 - 2003



(b) Promedio de la participación en el ingreso nacional del 10% más rico, 2015 - 2018

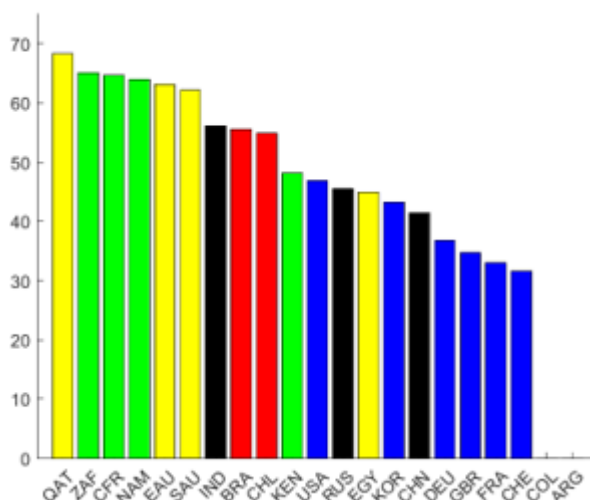


Figura A1. Participación del 10 % más rico en diversos países del mundo. Adaptado de WID.world. Elegimos algunos países de la base de WID.World; sus nombres son abreviados a su código internacional de tres letras. Las barras de países desarrollados (miembros de la OCDE) están rellenas en azul, y los demás países según región: América Latina (rojo), Asia (negro), Medio Oriente (amarillo) y África (verde). Todas las estadísticas han sido corregidas según alguna variante de la metodología DINA (con diferentes niveles de sofisticación, dependiendo del acceso a datos en ese país); los estudios de donde se WID.World recopilan estos datos están citados en la discusión de la metodología DINA del Capítulo II. En general, la definición de ingreso corresponde al ingreso nacional repartido a nivel de individuos, en vez de hogares, y antes de impuestos.

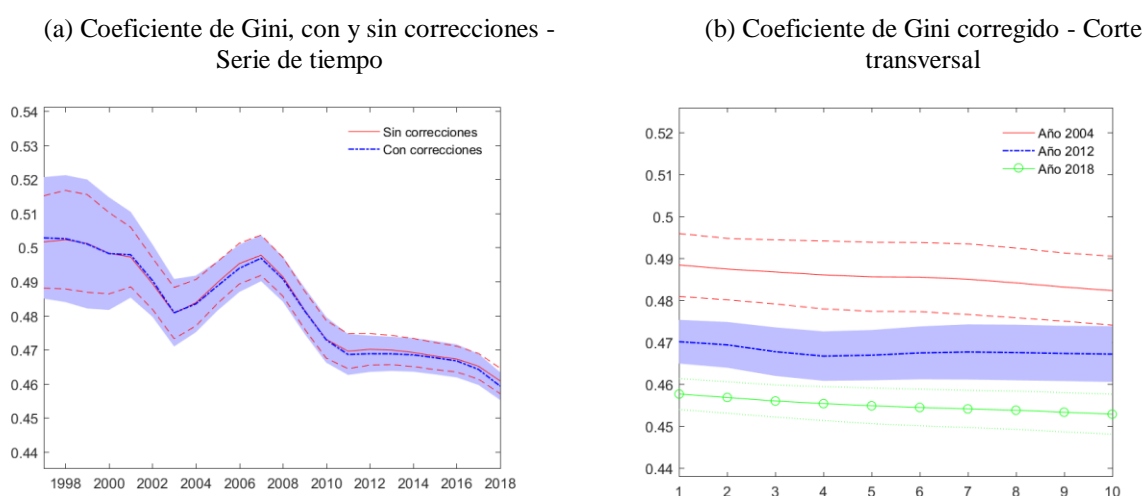


Figura A2. Estadísticas corregidas de la desigualdad del ingreso neto anual. Adaptado de Enaho 1997-2018.

La variable analizada en ambos paneles es el ingreso neto anual del hogar (inghog2d). El panel (a) y el panel (b) son los mismos cálculos que la Figura 5a y la Figura 6, respectivamente, pero aplicados a esta nueva variable. Para

clasificar, un hogar está en el grupo de mayores ingresos si se encuentra en el 2 % más rico de la distribución de ingresos (los resultados cualitativos se mantienen si se cambia el valor de este umbral). El error estándar del Gini no paramétrico es calculado por un *bootstrap* estándar; el error estándar del Gini semiparamétrico, por el *bootstrap* descrito en la metodología. Las series han sido suavizadas con un promedio móvil simétrico de tres años.

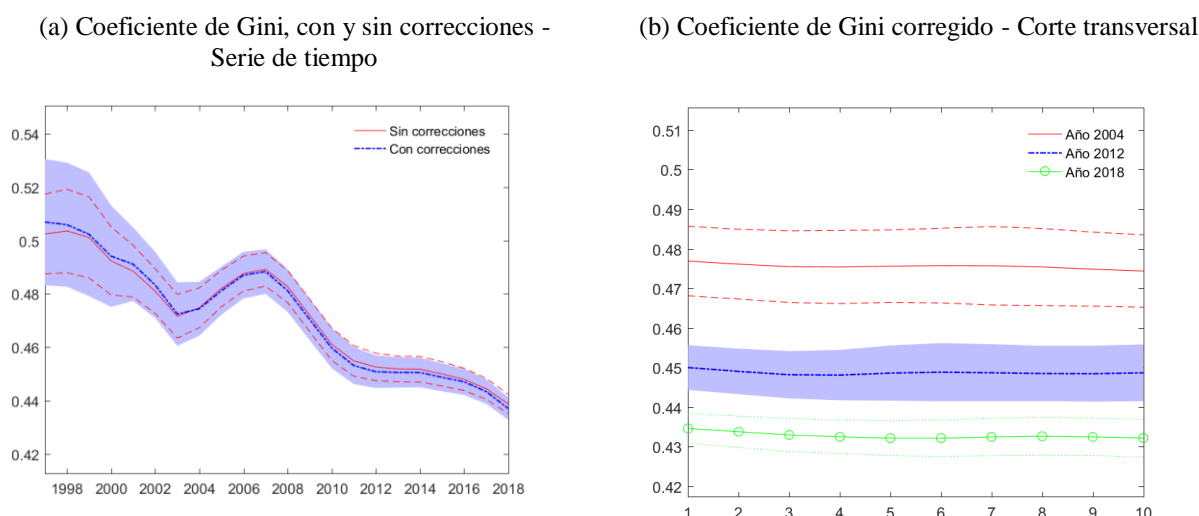


Figura A3. Estadísticas corregidas de la desigualdad del ingreso neto equivalente de los hogares. Adaptado de Enaho 1997-2018. La variable analizada en ambos paneles es el Ingreso Neto Anual del hogar (inghog2d) dividido entre la raíz cuadrada del número de miembros del hogar ($\text{mieperho}^{0.5}$) para controlar por efectos de escala. El panel (a) y el panel (b) son los mismos cálculos que la Figura 3a y la Figura 4, respectivamente, pero aplicados a esta nueva variable. Para clasificar, un hogar está en el grupo de mayores ingresos si se encuentra en el 2 % más rico de la distribución de ingresos (los resultados cualitativos se mantienen si se cambia el valor de este umbral). El error estándar del Gini no paramétrico es calculado por un *bootstrap* estándar; el error estándar del Gini semiparamétrico, por el *bootstrap* descrito en la metodología. Las series han sido suavizadas con un promedio móvil simétrico de 3 años.

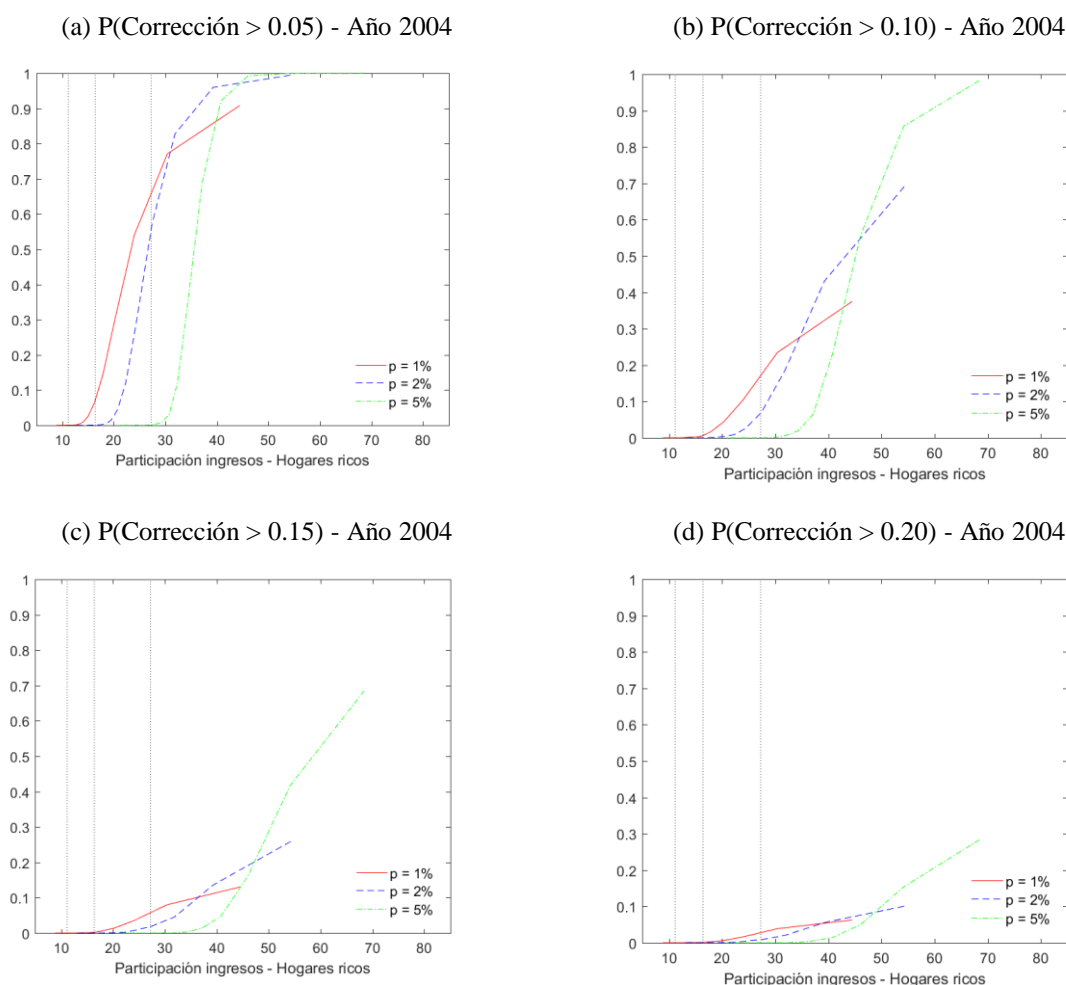


Figura A4. Probabilidades simuladas de observar correcciones de diversa magnitud en el coeficiente de Gini del año 2004. Adaptado de Enaho 2004. La variable analizada es el ingreso neto anual del hogar, dividido entre el número de miembros del hogar. El eje vertical indica la probabilidad simulada de observar que el Gini bajo los supuestos de esa simulación sea mayor que el Gini simulado con los estimadores de la muestra. La probabilidad es la proporción de veces que el evento de interés (observar un Gini hipotético mayor al de la muestra) ocurre en 1,000 repeticiones del algoritmo descrito en el Capítulo III. El eje horizontal y el color de cada línea indican la participación en el ingreso de los hogares ricos y la proporción de hogares no observados asumidos para el cálculo de esa probabilidad, respectivamente. En la programación probamos valores de $p = \{1, 2, 5\}$ y varios valores del índice de cola α desde 1.1 hasta 4; dados estos valores y la muestra, la distribución está completamente especificada y podemos resolver la participación de ingreso del p % de hogares con mayores ingresos s . Las líneas verticales punteadas señalan la participación del 1 %, 2 % y 5 % de hogares con mayores ingresos según los datos de la Enaho como referencia.

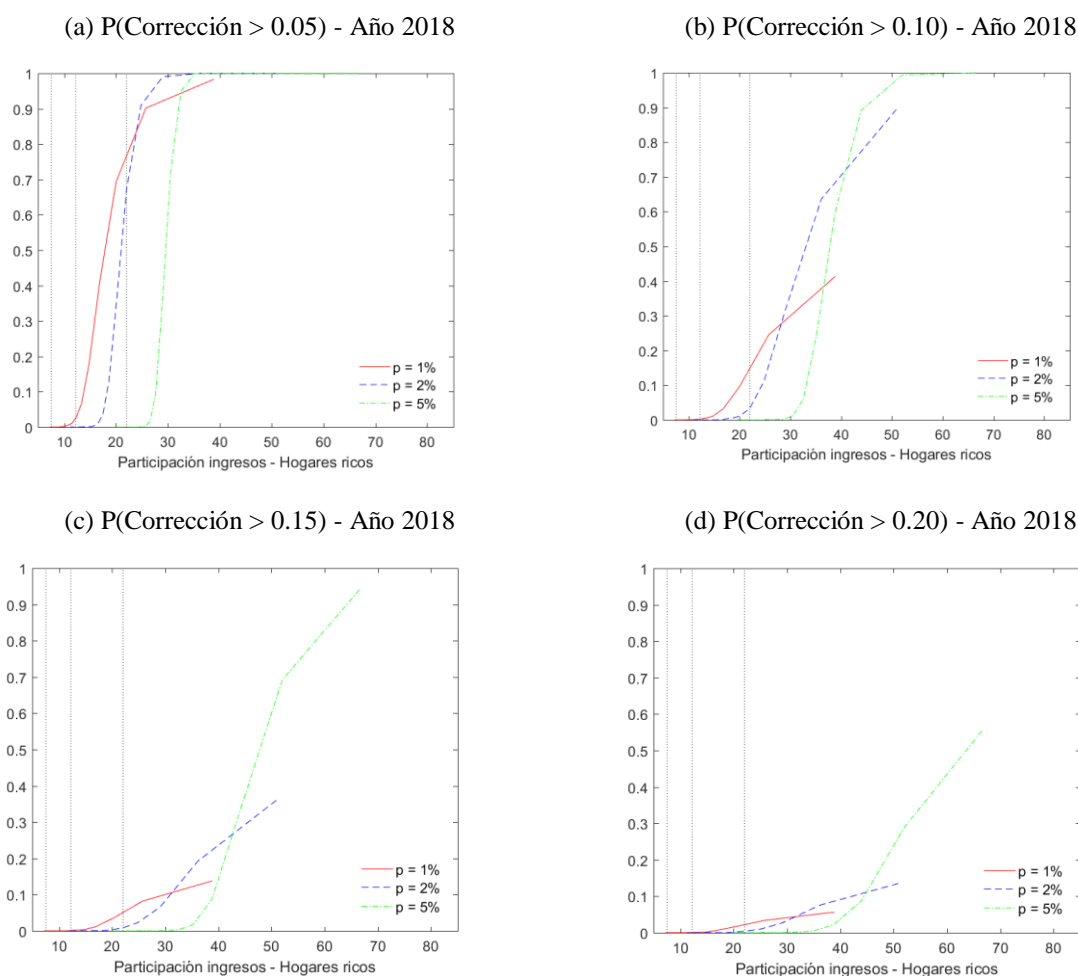


Figura A5. Probabilidades simuladas de observar correcciones de diversa magnitud en el coeficiente de Gini del año 2018. Adaptado de Enaho 2018. La variable analizada es el ingreso neto anual del hogar, dividido entre el número de miembros del hogar. El eje vertical indica la probabilidad simulada de observar que el Gini bajo los supuestos de esa simulación sea mayor que el Gini simulado con los estimadores de la muestra. La probabilidad es la proporción de veces que el evento de interés (observar un Gini hipotético mayor al de la muestra) ocurre en 1000 repeticiones del algoritmo descrito en el Capítulo III. El eje horizontal y el color de cada línea indican la participación en el ingreso de los hogares ricos y la proporción de hogares no observados asumidos para el cálculo de esa probabilidad, respectivamente. En la programación probamos valores de $p = \{1, 2, 5\}$ y varios valores del índice de cola α desde 1.1 hasta 4; dados estos valores y la muestra, la distribución está completamente especificada y podemos resolver la participación de ingreso del p % de hogares con mayores ingresos s . Las líneas verticales punteadas señalan la participación del 1 %, 2 % y 5 % de hogares con mayores ingresos según los datos de la Enaho como referencia.

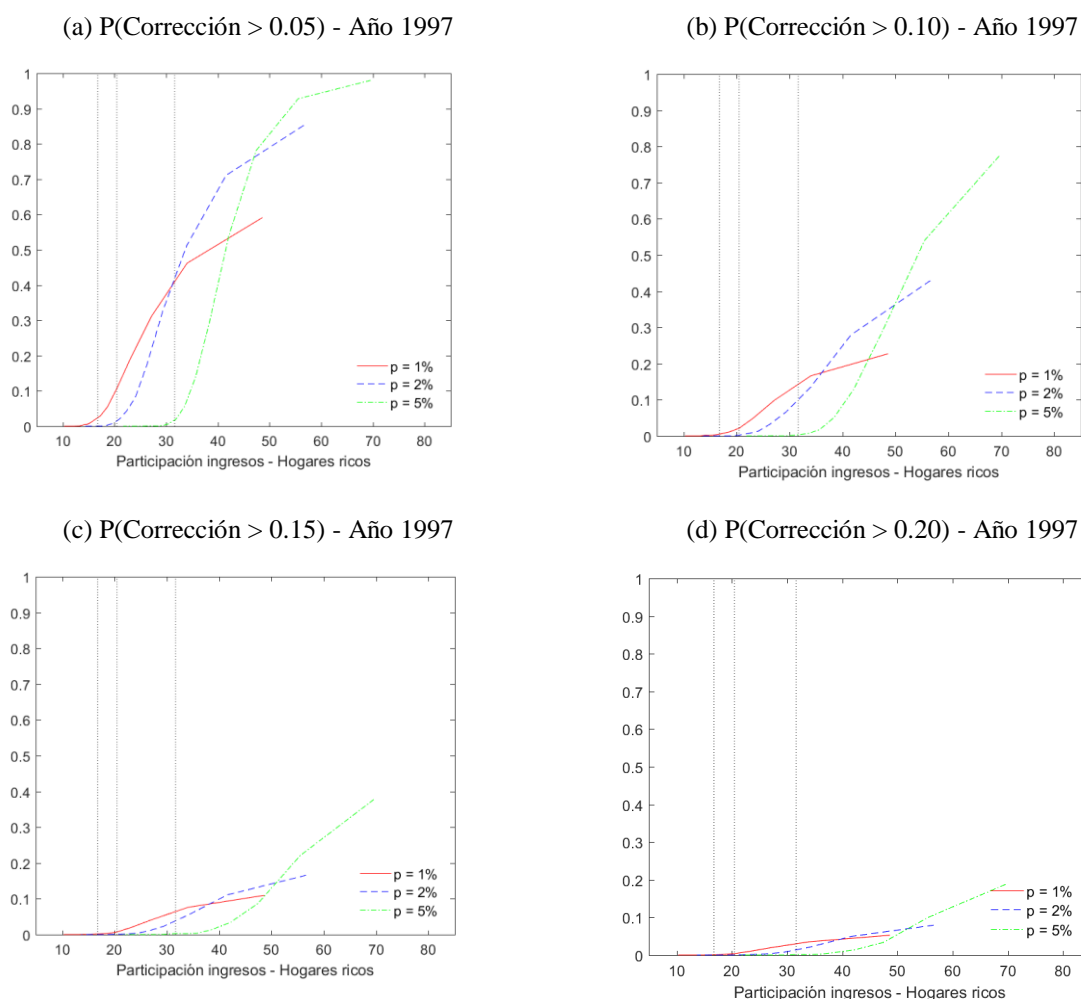


Figura A6. Probabilidades simuladas de observar correcciones de diversa magnitud en el coeficiente de Gini del año 1997. Adaptado Enaho 1997. La variable analizada es el ingreso neto anual del hogar, dividido entre el número de miembros del hogar. El eje vertical indica la probabilidad simulada de observar que el Gini bajo los supuestos de esa simulación sea mayor que el Gini simulado con los estimadores de la muestra. La probabilidad es la proporción de veces que el evento de interés (observar un Gini hipotético mayor al de la muestra) ocurre en 1000 repeticiones del algoritmo descrito en la sección III.C. El eje horizontal y el color de cada línea indican la participación en el ingreso de los hogares ricos y la proporción de hogares no observados asumidos para el cálculo de esa probabilidad, respectivamente. En la programación probamos valores de $p = \{1, 2, 5\}$ y varios valores del índice de cola α desde 1.1 hasta 4; dados estos valores y la muestra, la distribución está completamente especificada y podemos resolver la participación de ingreso del p % de hogares con mayores ingresos s . Las líneas verticales punteadas señalan la participación del 1 %, 2 % y 5 % de hogares con mayores ingresos según los datos de la Enaho como referencia.

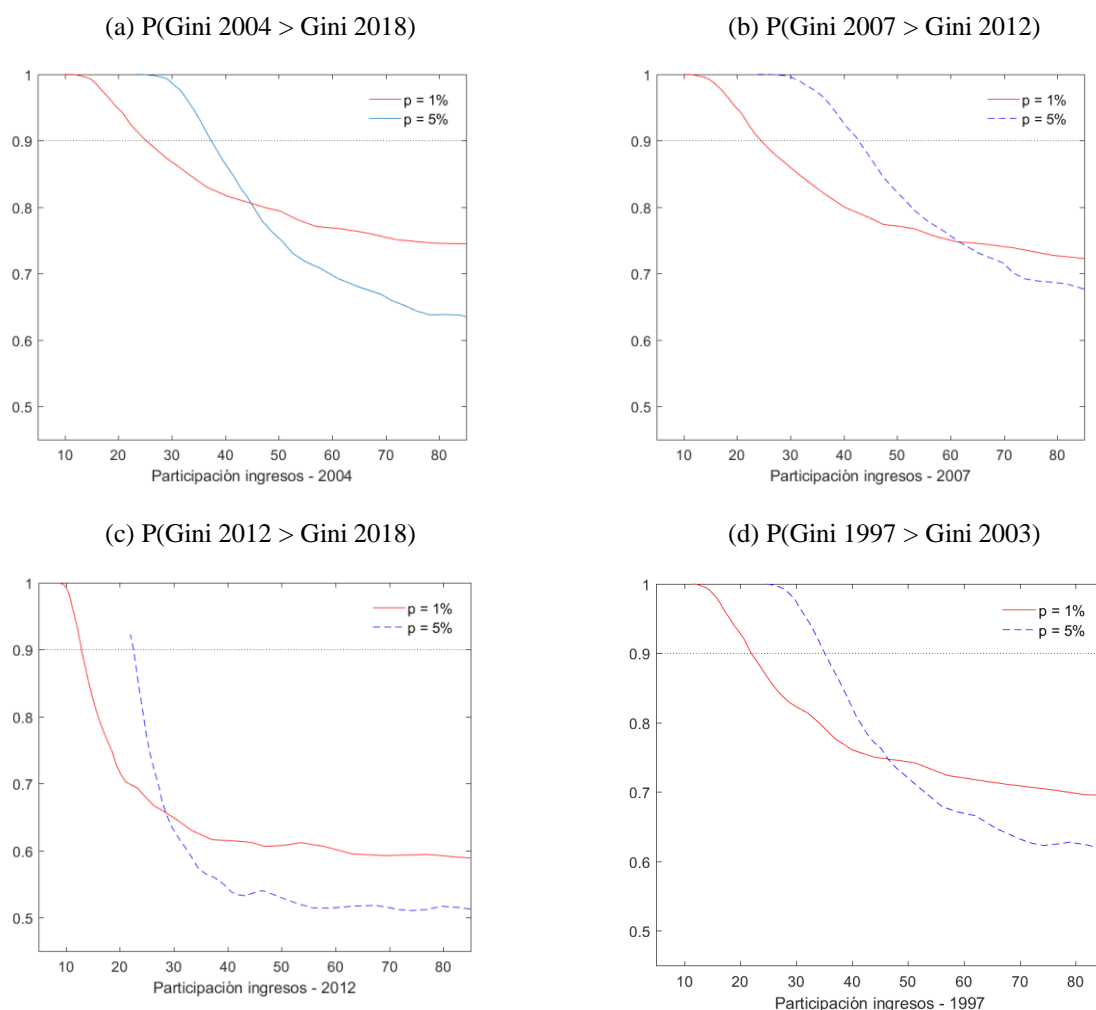


Figura A7. Probabilidades simuladas de observar cambios del Gini del ingreso bruto anual per cápita para varios años. Adaptado de Enaho 1997-2018. La variable analizada es el ingreso bruto anual del hogar (antes de impuestos), dividido entre el número de miembros del hogar. El eje vertical indica la probabilidad simulada de observar que el Gini en el primer año sea mayor que el Gini en el segundo año. La probabilidad es la proporción de veces que este evento de interés ocurra en 5,000 repeticiones del algoritmo descrito en el Capítulo III. El eje horizontal y el color de cada línea indican la participación en el ingreso de los hogares ricos del primer año y la proporción de hogares no observados asumidos para el cálculo de esa probabilidad, respectivamente. Si bien no presentamos la participación asumida para el segundo año, en todas las simulaciones, fijamos el mismo índice de cola para simular el Gini de cada año, y luego resolvemos el nivel de participación en el primer año; de esa forma también se podría resolver el nivel de participación del segundo año. Los valores del índice de cola α que probamos van desde 1.01 hasta 3.